

## Montclair State University Montclair State University Digital Commons

Department of Psychology Faculty Scholarship and  
Creative Works

Department of Psychology

Spring 4-2008

# Cross-modal interaction between vision and hearing: A speed—accuracy analysis


Yoav ArieH

Montclair State University, [ariehy@montclair.edu](mailto:ariehy@montclair.edu)

Lawrence E. Marks

Yale University, [lawrence.marks@yale.edu](mailto:lawrence.marks@yale.edu)

Follow this and additional works at: <https://digitalcommons.montclair.edu/psychology-facpubs>

 Part of the [Behavioral Neurobiology Commons](#), [Clinical Psychology Commons](#), [Cognitive Neuroscience Commons](#), [Health Psychology Commons](#), [Human Factors Psychology Commons](#), [Nervous System Commons](#), [Other Neuroscience and Neurobiology Commons](#), [Sense Organs Commons](#), [Speech and Hearing Science Commons](#), [Speech Pathology and Audiology Commons](#), and the [Systems Neuroscience Commons](#)

### MSU Digital Commons Citation

ArieH, Yoav and Marks, Lawrence E., "Cross-modal interaction between vision and hearing: A speed—accuracy analysis" (2008). *Department of Psychology Faculty Scholarship and Creative Works*. 26.  
<https://digitalcommons.montclair.edu/psychology-facpubs/26>

### Published Citation

ArieH, Yoav, and Lawrence E. Marks. "Cross-modal interaction between vision and hearing: A speed—accuracy analysis." *Perception & Psychophysics* 70, no. 3 (2008): 412-421.

# Cross-modal interaction between vision and hearing: A speed–accuracy analysis

YOAV ARIEH

*Montclair State University, Montclair, New Jersey*

AND

LAWRENCE E. MARKS

*John B. Pierce Laboratory, New Haven, Connecticut  
and Yale University, New Haven, Connecticut*

Cross-modal facilitation of response time (RT) is said to occur in a selective attention task when the introduction of an irrelevant sound increases the speed at which visual stimuli are detected and identified. To investigate the source of the facilitation in RT, we asked participants to rapidly identify the color of lights in the quiet and when accompanied by a pulse of noise. The resulting measures of accuracy and RT were used to derive speed–accuracy trade-off functions (SATFs) separately for the noise and the no-noise conditions. The two resulting SATFs have similar slopes and intercepts and, thus, can be treated as overlapping segments of a single function. That speeded identification of color with and without the presence of noise can be described by one SATF suggests, in turn, that cross-modal facilitation of RT represents a change in decision criterion induced by the auditory stimulus. Analogous changes in decision criteria might also underlie other measures of cross-modal interactions, such as auditory enhancement of brightness judgments.

Multisensory research is a rapidly growing field that considers the ways in which different senses interact (Calvert, Spence, & Stein, 2004). These interactions are deemed crucial to the elaborate process of building a coherent picture of the environment. For example, visual and auditory cues serve together to help pinpoint the source of incoming sounds. When these cues are dissociated, gross errors may occur in sound localization (the ventriloquist effect; Alais & Burr, 2004; Jack & Thurlow, 1973).

One of the basic phenomena that emerged from early multisensory research is cross-modal facilitation of response time (RT) in a selective attention task (Nickerson, 1973; Todd, 1912). When people are asked to respond to a visual stimulus—say, to determine the lateral position of a flash of light—their choice RT (CRT) is likely to be shorter if the visual stimulus is accompanied by an irrelevant auditory signal (Bernstein, Clark, & Edelman, 1969; Bernstein & Edelman, 1971; Simon & Craft, 1970). This result is surprising because the auditory signal, by definition, carries no information relevant to the visual choice. Nevertheless, the effect is robust and has been replicated under diverse experimental settings. What mechanism might account for the cross-modal facilitation of visual CRT?

Two general classes of mechanisms have been proposed to account for the facilitation, one class being *sensory* and the other *decisional*. Sensory mechanisms include

prominently the hypothesis of energy summation (or integration): that some of the energy in the auditory stimulus combines with the energy in the visual stimulus at a relatively early stage of information processing, thereby increasing the effective visual intensity (Bernstein et al., 1969; Bernstein, Rose, & Ashe, 1970). In other words, the effect of the irrelevant acoustic signal is comparable, by this hypothesis, to that of increasing the luminance of the visual stimulus. In the language of signal detection theory, this account implies that cross-modal facilitation in CRT occurs because of a shift in visual *sensitivity* wrought by adding acoustic energy. Other sensory mechanisms are also possible. For example, the addition of an auditory stimulus could serve as a signal that reduces temporal uncertainty, thereby increasing the ratio of signal to noise in the visual system and, hence, visual sensitivity.

Alternatively, cross-modal facilitation of RT might reflect the operation of decisional processes. An example is the preparation enhancement hypothesis (Nickerson, 1973), which suggests that the effect of the auditory signal is essentially the same as that of a warning signal that prompts the observer to respond (see also Posner, Nissen, & Klein, 1976). Rather than increasing the detectability or discriminability of the visual target, however, this hypothesis states that the auditory stimulus causes the observers to rely on less information when making their

---

Y. Arieh, [ariehy@mail.montclair.edu](mailto:ariehy@mail.montclair.edu)

---

decision and, thus, to respond sooner. In the language of signal detection theory, this hypothesis implies that cross-modal facilitation in CRT reflects a shift in the *decision criterion* that observers use to choose when to execute a particular response.

The purpose of the present study is to choose critically between sensory and decisional accounts of cross-modal facilitation in tasks of selective attention to visual stimuli, as measured by CRT. Our general strategy is to compare two speed-accuracy trade-off functions (SATFs) derived from CRTs made to visual stimuli (visual CRTs, for short). One SATF represents performance in the presence of an irrelevant auditory noise, and the other represents performance in the quiet. The SATF depicts how accuracy varies with time (speed) of response and can be characterized by two parameters: slope and intercept. Any difference in either the slope or the intercept between the SATFs obtained with and without noise implies that the irrelevant noise affected visual sensitivity, as well as, perhaps, criterion, whereas comparable values of both the slope and the intercept would imply that the noise affected only the decisional criterion. But first, we will briefly review the empirical evidence at hand regarding cross-modal interactions using several measures of performance, evidence that does not allow a clear choice between these alternatives. Then, we will elaborate on the logic of the SATF methodology and describe its implementation in the present study.

### **Cross-Modal Facilitation of CRT: Sensitivity Increase or Decisional Shift?**

Over the past 2 decades, there has been an enormous increase in research on multisensory integration and interaction, a good deal of it focusing on integration and interaction between vision and hearing (see the volume edited by Calvert et al., 2004). Furthermore, much of this work has sought evidence, or assumed, that multisensory integration and interaction in perception, including cross-modal facilitation in tasks of selective attention, largely represent the outcome of multisensory integration and interaction in the sensory nervous system. In general, sensory explanations of cross-modal facilitation imply that changes in sensitivity occur even when performance is assessed with decision-free or decision-controlled methods. If multisensory integration and interaction are widespread properties of sensory processing, evidence of energy summation, for example, might be evident in a variety of perceptual measures, not just in tasks of selective attention or measures of RT.

A straightforward consequence of the energy summation hypothesis, for example, is that the detection threshold of a combination of sound and light (bimodal detection) should be smaller than the threshold of either stimulus alone (unimodal detection). Findings to date, however, have not offered clear support for this prediction. In some studies, the small gains in bimodal detection that have been reported can either be explained by models of probability summation that assume sensory independence (Brown & Hopkins, 1967; Mulligan & Shaw, 1980) or be eliminated when criterion-controlled methods are used (Loveless, Brebner, & Hamilton, 1970). A few studies in

which suprathreshold auditory stimuli have accompanied threshold-level visual test stimuli (tasks of selective attention at threshold) have reported small gains in visual sensitivity (as measured by  $d'$ ) when the sound and the light were presented simultaneously at the same location (Bolognini, Frassinetti, Serino, & Ládavas, 2005; Frassinetti, Bolognini, & Ládavas, 2002). Interestingly, these studies have also reported changes in response criteria (as measured by  $\beta$ ) that were independent of the degree of spatial or temporal coincidence of the lights and sounds. That is, besides any increase in sensitivity, the presence of a perceptible acoustic stimulus increased the observers' willingness to report a visual stimulus.

If an acoustic stimulus can increase the effective energy in the visual neural pathways, another reasonable prediction is that a visual stimulus will appear brighter in the presence of the irrelevant auditory stimulus than in its absence, as Stein, London, Wilkinson, and Price (1996) reported: Observers rated weak flashes of light to be brighter when the flashes were accompanied by a concurrent pulse of white noise (see also Odgaard, Arieh, & Marks, 2003, Experiment 1, for a replication). Stein et al. proposed that enhancement occurs because energy from the visual and the auditory pathways is integrated to augment brightness, speculating that this integration may take place in the superior colliculus, where multimodal neurons reach peak firing rate only in the presence of spatial-temporal congruent auditory and visual inputs (Stein & Meredith, 1993; see also Meredith, 2002, for a review). As others have noted, however, the brightness ratings in Stein et al. cannot unambiguously distinguish between sensory and decisional change (Vroomen & de Gelder, 2000). The presence of irrelevant noise could have led the observers to shift their rating scale, relative to the underlying representations of brightness, so that less light energy was needed to elicit a given rating, a plausible explanation in terms of decisional mechanisms (see also Odgaard et al., 2003). Thus, neither the findings on detection threshold nor those on suprathreshold brightness provide clear-cut evidence for energy summation or for other sensory explanations of the reported interactions.

The decisional shift hypothesis specifies the relation between speed and accuracy in a task requiring rapid visual choice, when an irrelevant auditory stimulus is presented: Because the irrelevant sound reduces the amount of information that is being used to choose the appropriate response, gains in speed will be offset by losses in accuracy. In other words, the irrelevant sound produces a speed-accuracy trade-off in visual CRT. Indeed, it is well documented that participants commit more errors in the presence of irrelevant sounds than in their absence (Nickerson, 1973; Posner, 1978; Posner et al., 1976; Schmidt, Gielen, & van den Heuvel, 1984). The data reported by Schmidt et al. (Experiment 1) are typical: In the presence of irrelevant acoustic signals, participants performed the visual choice task more quickly by 30 msec but their accuracy dropped from about 99% to 94.75%. Unfortunately, these data by themselves cannot distinguish between the sensory and decisional hypotheses. First, to explain the facilitation solely by a speed-accuracy trade-off mecha-

nism, one would have to show that the decline in RT of 30 msec is equivalent to the decrease in accuracy from 99% to 94.75%. However, the trade-off between speed and accuracy can be recovered only from an SATF, and to the best of our knowledge, the present study is the first to compute the speed–accuracy conversion in the context of cross-modal facilitation of CRT. Second, the overall accuracy rates reported by Schmidt et al. and in other related studies were relatively high (e.g., above 90%). In fact, as will be discussed below, these levels of accuracy reside in a region in which the SATF usually asymptotes and, thus, does not provide useful information about the relation between speed and accuracy in selective attention tasks measuring CRT.

### The Speed–Accuracy Trade-Off Function in CRT

The SATF is a well-recognized analytical tool that has often been used to characterize the dynamics of information processing over time in diverse fields, such as suppression of loudness by transients (Arieh & Marks, 2003), visual acuity across the retina (Carrasco, McElree, Denisova, & Giordano, 2003), working memory (McElree, 2001), and the effects of stimulus uncertainty (Lappin & Disch, 1972a). An SATF can be derived readily from data obtained in a speeded choice task because the task provides measures of errors, as well as RT, and thus makes it possible to assess relative changes in these two measures across experimental manipulations. When RT and errors in a choice task are manipulated by enforcing longer and shorter response deadlines and then accuracy is plotted against RT, the result is a monotonically increasing, negatively accelerated function. In other words, the typical SATF shows that up to some asymptotic level, accuracy increases with processing time (Pachella, 1974; Pew, 1969). An SATF can be produced in two ways. One way is to manipulate speed and accuracy across experimental conditions—for example, by enforcing different deadlines, as just described. Another way to produce an SATF is to exploit the natural variability in speed and accuracy that arises from trial to trial within a single session—essentially, to determine how errors within a session decrease as RT increases. To distinguish these approaches, some authors have labeled the function obtained through the second approach as a conditional accuracy function (e.g., Ollman, 1977). Here, however, for the sake of clarity in presentation and because our goal is to relate speed to accuracy, we will use the generic term SATF.

Many models of CRT assume that the stimulus produces a sensory effect that drifts noisily over time toward criteria or thresholds. Sequential-sampling models of choice response, such as evidence accumulation and random walk models, differ in their details but agree in that they interpret different points along a given SATF as reflecting shifts in response criteria or boundaries, and not as changes in the sensory representation of the stimulus (see Luce, 1986, for a thorough review). Presumably, when a person is induced to respond quickly, the criterion or boundary for responding is set at a relatively low value. Because the criterion is low, less information is allowed to accumulate before the response is made, and because the

accumulation of information over time is a noisy process, accuracy, as well as RT, declines. Alternatively, raising the criterion allows more information to accumulate, so that both accuracy and RT increase. Given this framework, the SATF can be used readily to test hypotheses about possible sources of cross-modal interaction in tasks of speeded choice. For example, if presenting an irrelevant sound induces only a shift in criterion in a visual CRT task, the joint measures of speed and accuracy obtained with and without the sound should lie on a single SATF.

For convenience of exposition, we shall consider SATFs that plot accuracy (e.g.,  $d'$ ) as a function of RT (in milliseconds), so that prototypical functions will increase (accuracy increases as processing time increases). An SATF is generally characterized by its slope, intercept, and asymptote (Doshier, 1979; McElree, 2001; Wickelgren, 1977) or by its slope and intercept, if one considers only the region before the function asymptotes (Lappin & Disch, 1972a, 1972b, 1973). The intercept (on the axis of time) is interpreted as a temporal threshold that reflects the minimum processing time necessary for performance to begin rising above chance. In other words, the intercept is the point in time beyond which sufficient information has accumulated so that speed can begin to trade with accuracy. The slope is interpreted as the rate at which information accumulates over time. The faster the accumulation of evidence about the feature of the stimulus pertinent to the choice task, the greater the extent to which accuracy increases per unit of processing time and, thus, the steeper the slope of the SATF.

In the framework of a speed–accuracy analysis, an irrelevant sound might speed the choice response to visual stimuli in two ways. The sound might induce the observer to lower the decisional criterion that governs the amount of information that must accumulate before a response is made, leading to short RT but lower accuracy—consistent with preparation enhancement and other decision-based hypotheses of facilitation. Alternatively, or in addition, and consistent with the energy summation hypothesis, the sound might add some of its energy to that of the light. An increase in effective signal strength could reveal itself as either (1) a change in intercept, a lower temporal threshold beyond which accuracy increases beyond chance level, or (2) a change in slope, a greater trade-off ratio of accuracy for time ( $d'/\text{msec}$ ).

Note, in this regard, that in a visual choice task, Lappin and Disch (1972b) reported that increasing the luminance of a visual stimulus both increased the slope and decreased the intercept of the resulting SATF. This finding is especially pertinent to the present study and to evaluating the hypothesis that auditory facilitation of visual CRT results from summation of sound and light energy. Given the findings of Lappin and Disch (1972b), the energy summation hypothesis implies that adding an irrelevant, facilitating sound should both increase the slope and decrease the intercept of the visual SATF.

### The Present Study

In this selective attention paradigm, we asked participants to identify the colors green and red in two experimental conditions: in the presence of an irrelevant pulse

of noise (noise-on condition) and in the quiet (noise-off condition). To construct a reliable SATF for each condition, we used a response deadline procedure that induced the participants to respond quickly and, thus, to commit a substantial number of errors. This procedure was necessary because the focus of this study was the region of the SATF at which accuracy grows with RT, rather than its asymptotic region. Assuming overall auditory facilitation of visual CRT, Figure 1 depicts the four possible relations between SATFs in the noise-on and noise-off conditions. Panels A, B, and C show that the SATF for the noise-on condition (dashed line) can have a steeper slope (A), a smaller intercept (B), or both (C), relative to the SATF in the noise-off condition (solid line). In each of these cases, information reaches criterion (represented by the light faced horizontal line) sooner in the noise-on than in the noise-off condition, thereby producing facilitation.

Changes in slope of the SATF (panels A and C) are consistent with the hypothesis of energy summation: Let  $I_v$  represent the effective intensity of the visual stimulus and  $I_a$  the portion of the effective intensity of the auditory stimulus that putatively adds to visual intensity. In a simple model of information accumulation, the rate of information accumulation (slope of the SATF) is related directly to the overall intensity by a function,  $F$ —the brighter the stimulus, the faster the accumulation and the steeper the slope (Pins & Bonnet, 1996). Thus, the slope would be proportional to  $F(I_v)$  when the light is presented alone but proportional to  $F(I_v + I_a)$  when the light is accompanied by an irrelevant sound. According to the energy summation hypothesis, therefore, the increase in slope of the SATF is directly related to the contribution made by the intensity of the auditory stimulus.

Changes in the intercept of the SATF (panels B and C) need not, strictly speaking, follow from a model of energy summation that assumes an accumulation of intensity-dependent information from both modalities, although the data of Lappin and Disch (1972b) indicate that changes in luminance can modify the intercept, as well as the slope, of visual SATFs. A model of cross-modal energy summation can predict changes in intercept, as well as changes in slope (panel C), if one makes the Fechnerian assumptions that subthreshold energy summation begins at a constant

time after the initiation of the stimulus (assuming a constant neural latency) and that, in addition to a presumably constant time to execute the motor response, the intercept reflects a threshold that must be reached before accuracy rises from a chance level—that is, before  $d'$  exceeds zero. Changes in intercept are also consistent, however, with other possible sensory models of cross-modal facilitation. One plausible model that accounts for a change in intercept without a change in slope (panel B) rests on the evidence that auditory information is processed more quickly than visual information (Hershenson, 1962; see Woodworth & Schlosberg, 1954). Perhaps an irrelevant auditory stimulus reduces visual RTs by “priming” the visual system to start accumulating the information relevant to the response sooner than the system would accumulate information when the auditory stimulus is absent, but without an ancillary change in rate of accumulation. The result would be a decrease in the intercept without a change in the slope of the SATF.

Regardless of the specific mechanism, panels A–C of Figure 1 show three possible outcomes that are consistent with energy summation and other sensory mechanisms of cross-modal facilitation. Panel D, by contrast, assumes that an irrelevant sound affects only the location of the response criteria, without affecting the rate by which information accrues, and thus predicts that a single SATF will characterize performance measured with and without the irrelevant acoustic stimulus.

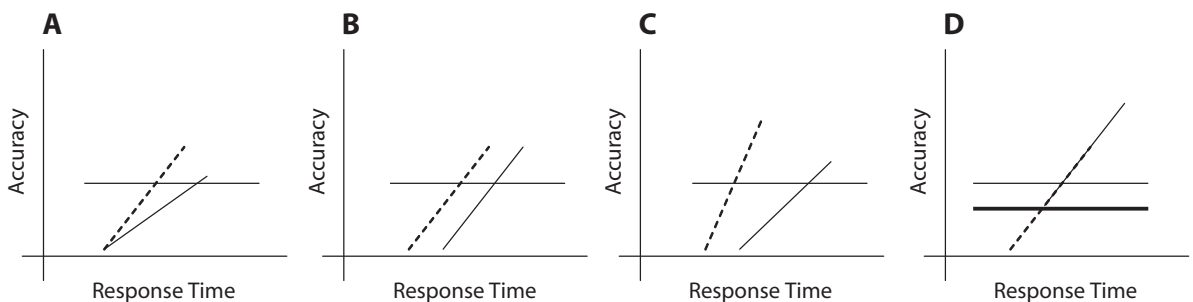
## METHOD

### Participants

All the participants were recruited from the Montclair State University community and self-reported having normal hearing and normal vision. Fifteen women and 3 men received course credit for participating in the two sessions.

### Apparatus and Stimuli

**Visual stimuli.** The flashes of green and red light were circular disks, 0.2 cm in diameter, digitally produced by a NVIDIA GeForce4 video card that was controlled through the Psychophysical Toolbox for MATLAB (Brainard, 1997) and housed in a Pentium 4 Dell computer. Flashes were presented on a black background at the center of a 17-in. Dell flat monitor with a refresh rate of 70 Hz for four frames, thus fixing their duration at 57 msec. The RGB values were



**Figure 1.** Four possible relations between the speed-accuracy trade-off functions (SATFs) of the noise-on (dashed line) and the noise-off (solid line) conditions. Panels A–C show changes in slope, in intercept, and in both, respectively. Panel D, however, shows similar slopes and intercepts for the noise-on and the noise-off conditions, resulting in overlapping SATFs. The lightfaced horizontal line represents the decision criterion, and the boldfaced vertical line in panel D represents a lower position of the decision criterion.

[0 50 0] and [50 0 0] for the green and the red disks, respectively. The visual stimuli produced an illuminance of 0.5 ft-c when a digital light meter (Model EA30, Extech Instruments, Waltham, MA) was placed directly at the surface of the screen. The participants viewed the light flashes from a distance of 57 cm through a chinrest that served to fix the position of their head.

**Auditory stimuli.** The auditory stimuli were 57-msec pulses (rise and fall of 5 msec) of broadband random noise with a frequency range of 20 Hz–20 kHz. The signal was produced by a Tucker-Davis Technologies System 3 (Alachua, FL) and then fed to a Blaupunkt PCx 352 loudspeaker positioned directly above the screen. The level of the noise pulse was set to 50 dB SPL (A), measured at the chinrest with Brüel & Kjaer Type 2240 sound-level meter.

The Tucker-Davis System also controlled the measurement of RT via a Tucker-Davis RB-25 response box with a sampling frequency of 50 kHz, which supplied a dedicated clock with better-than-millisecond accuracy. All aspects of stimulus presentation, randomization, data collection, and data recording were handled by a dedicated MATLAB program.

We manipulated the presence of the auditory stimulus to create two conditions: (1) a *noise-on* condition, in which the burst of noise was presented simultaneously with the colored disks (to ensure auditory–visual simultaneity, the timing of auditory presentation was controlled by an external Tucker-Davis clock that was synchronized to the PC video card) and (2) a *noise-off* condition, in which the colored disks were presented without the noise. Combining these two conditions (noise-on, noise-off) with the two colors (red, green) resulted in the four basic experimental stimuli in this study. The four stimuli were intermixed randomly within an experimental block that contained 12 repetitions of each condition. An experimental session contained seven blocks, the first of which was treated as practice and was excluded from the analyses.

### Procedure

The sessions were conducted in a dark, sound-attenuated booth. After 10 min of dark adaptation, the participants entered the booth and received the instructions. The participants were told that a series of red and green disks would be presented on the computer screen and that these disks would be accompanied by soft bursts of noise on some of the trials. The participants were also told that their task was to ignore the noise and identify each disk as red or green by pressing the appropriate button on the response box as quickly as they could without worrying about accuracy. The time interval between successive presentations of the disks was 1.5 sec. Because the main goal was to construct SATFs, we used a *deadline procedure* to ensure that the participants would both respond quickly and make a significant number of errors. We fixed the deadline at 450 msec, a value that produced around 75% accuracy in a short pilot study. Because 75% lies between perfect accuracy and chance performance, we used this value to target the region of the SATF at which the relation between speed and accuracy should be approximately linear. Each time an RT exceeded the deadline, a minus sign appeared in the center of the screen, cuing the participant to respond more quickly. Indeed, overall accuracy, across all participants and stimuli, was 78.5%, indicating that our deadline procedure was successful in maintaining the level of accuracy needed for producing informative SATFs. The participants took part in two sessions that were at least 24 h apart. Because each session contained 336 trials, there were 672 RTs per participant, or 168 RTs per experimental condition.

## RESULTS

### Response Time and Accuracy in the Noise-On and Noise-Off Conditions

First, we tested whether the presence of noise would produce (1) a facilitation of RT and (2) a reduction in accuracy. For each participant, we transformed percent correct to  $d'$ , using a 2 (stimulus: red or green)  $\times$  2 (response:

red or green) confusion matrix computed separately for each participant for the noise-on and the noise-off conditions (Lappin & Disch, 1972a; Swensson, 1972). *Hits* were defined as proportions of *red* responses to red stimuli, whereas *false alarms* were defined as proportions of *red* responses to green stimuli. The resulting values of  $d'$  were then averaged across participants. For each condition, we also calculated the overall median RT, pooling RTs to both correct and incorrect responses, and averaged the median values across participants. The averages of  $d'$  and RT are presented in Figure 2.

The participants identified the color as red or green significantly more quickly when irrelevant noise was present, rather than absent. Average RT was 349 msec in the noise-on condition and 389 msec in the noise-off condition. The difference of 40 msec is statistically significant [ $t(17) = 7.3, p < .05$ ]. On the other hand, the participants identified the color more accurately when the noise was absent, rather than present. Average  $d'$  was 1.42 and 1.92 for the noise-on and the noise-off conditions, respectively. The difference of 0.5 in  $d'$  is also statistically significant [ $t(17) = 6.6, p < .05$ ].

Thus, our results agree with many reports that showed faster responses but reduced accuracy when people identified visual stimuli in the presence of irrelevant auditory stimulation (Nickerson, 1973; Posner et al., 1976). At first pass, the results suggest a speed–accuracy trade-off between the noise-on and the noise-off conditions: The participants were faster to decide between red and green in the noise-on condition than in the noise-off condition, but they “paid” for the added speed with a reduction in accuracy. For these results to be explained completely by a speed–accuracy trade-off mechanism, however, we must assume that a facilitation of 40 msec in RT is equivalent to a decline of 0.5  $d'$  unit in accuracy. To determine whether this is so, one must measure the SATFs themselves under the noise-on and noise-off conditions. If facilitation in RT reflects a speed–accuracy trade-off, the SATFs in the noise-on and noise-off conditions should overlap. In fact, assuming a linear relation between RT and  $d'$ , we predicted that the slopes of both functions, in units of  $d'$ /msec,

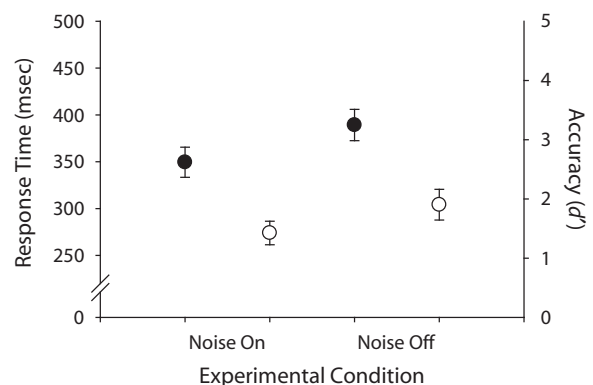


Figure 2. Average response time (filled circles) and accuracy ( $d'$ , empty circles) for the noise-on and the noise-off conditions. The error bars show one standard error of the mean.

should approximately equal  $0.5/40 = 0.0125$ , that participants would trade approximately 0.0125  $d'$  unit of poorer accuracy for each 1 msec in greater speed.

### Speed–Accuracy Trade-Off Analysis

The construction of SATFs for the noise-on and noise-off conditions followed closely the procedure outlined by Lappin and Disch (1972a, 1972b, 1973). This procedure, applied to data sets obtained from individual participants, takes advantage of the natural variability in the participants' responses. For each participant, we did the following: (1) rank ordered the RT distribution for each condition, (2) partitioned the ranked distribution into bins containing equal number of trials, (3) computed accuracy in  $d'$  units and median RT for each bin, (4) plotted  $d'$  against RT to create the SATF function, and, finally, (5) recovered the intercept and the slope of the SATF by regressing RT on  $d'$ .

Our choice of a linear function to describe the relation between RT and  $d'$  was empirically driven. First, although other functions have been suggested (such as an exponential; see McElree, 2001), Lappin and Disch (1972a) reported that a linear function characterized the speed–accuracy relation well if, as was also the case in the present study, the RTs do not approach asymptote. Second, visual inspection of the present SATFs revealed that most of them could be fitted reasonably well by a linear function.

In the present study, the participants identified 288 color stimuli in each of the two experimental conditions. To maximize the number of trials per bin and to optimize the resolution of the RT differences, we decided to divide the RT distribution into five bins (the first three bins had 58 trials, and the last two had 57 trials each). The value of  $d'$  for each bin was calculated by using the  $2 \times 2$  color  $\times$  response confusion matrix described earlier. In the few cases in which the percentage of false alarms was 0 or the percentage of hits was 1, we used the formula  $1/(2N)$  to correct the values of false alarms and the formula  $1 - 1/(2N)$  to correct the values of hits (where  $N$  is the maximum number of false alarms or maximum number of targets). Finally, the SATF for each condition was constructed by plotting the values of  $d'$  against the corresponding median values of RT.

In this manner, we constructed 18 SATFs for the noise-on condition and 18 SATFs for the noise-off condition. Visual inspection of the functions revealed that 30 of the 36 can be reasonably described by a straight line. The other 6 functions (4 in the noise-off condition and 2 in the noise-on condition) were linear throughout the first four bins, but accuracy dropped significantly at the last bin. The reason for the drop in accuracy at relatively long RTs in those 6 functions is presently unknown. We can only speculate that, on a fraction of the trials, some participants may have failed to attend adequately, perhaps inducing them to respond slowly on those trials and resort to guessing. Regardless of the exact reason, we decided to define the SATF function as extending up to the point of maximum accuracy.

To recover the slope and intercept for each individual SATF,  $d'$  was regressed linearly on RT to produce the best-fitting line, which thereby minimized deviations in accu-

racy. The results of the subject-by-subject regressions are presented in Table 1.

The summary line at the bottom of the table provides the most important feature of the data: On average, the slopes and the  $x$ -axis intercepts of the SATFs obtained in the noise-on and the noise-off conditions were almost identical. The slopes ranged from 0.007 to 0.021, with a mean of 0.0129 and a standard error of 0.001, in the noise-off condition and from 0.007 to 0.025, with a mean of 0.0128 and a standard error of 0.001, in the noise-on condition, [ $t(17) = 0.046$ , n.s.]. Thus, the presence of irrelevant noise did not change the rule of conversion between speed and accuracy in visual color identification. On average, 1 msec was worth about 0.013  $d'$  unit both when noise was present and when noise was absent, a value that is remarkably similar to the 0.0125  $d'$  unit per millisecond predicted from the overall mean facilitation in RT and cost in accuracy.

The presence or absence of irrelevant noise also had no appreciable effect on the  $x$ -axis intercept of the SATF. The intercept is often interpreted as the point in time at which performance exceeds chance (e.g.,  $d' > 0$ ). That is, the intercept defines a temporal threshold at which enough information has accumulated to surpass background noise. In keeping with this interpretation, we recovered the intercept in milliseconds from the original regression equation where the intercept was given in  $d'$  units. The intercepts ranged from 42 to 342 msec, with a mean of 201 and a standard error of 80, in the noise-off condition and from 192 to 256 msec, with a mean of 211 and a standard error of 31, in the noise-on condition. Note that the intercept was actually slightly larger in the noise-on condition than in the noise-off condition. Nevertheless, this small numerical difference between the means is not significant [ $t(17) = 0.5$ , n.s.].

Another way to appreciate the main finding of this study—that the SATFs obtained with and without an irrelevant noise are essentially the same—is to average over participants the values of RT and  $d'$  obtained for each bin and construct a pooled SATF for each condition. These pooled SATFs appear in Figure 3.

Clearly, the functions for the noise-on and noise-off conditions overlap. In fact, the functions conform to the pattern depicted in panel D of Figure 1. Essentially, the plot shows that speed–accuracy performance in the two experimental conditions can be described by one linear function. The regression on the averaged data surrendered values that are similar to those obtained in the subject-by-subject analysis. For the noise-on condition, the slope was 0.012 and the  $x$ -axis intercept was 203 msec (adjusted  $R^2 = 99\%$  and  $p < .05$ ), and for the noise-off condition, the slope was 0.012 and the  $x$ -axis intercept was 207 msec (adjusted  $R^2 = 99\%$  and  $p < .05$ ).

## DISCUSSION

The two main findings of this study are simple and clear-cut. First, an irrelevant noise decreases the time needed to identify the color of a light, overall RT being smaller in the presence versus the absence of noise. And second, this facilitation in RT can be fully explained in terms of a

**Table 1**  
**Slope (in  $d'$ /msec), Intercept (Recovered in Milliseconds**  
**From the Regression Equation), and Adjusted  $R^2$  for**  
**Individual Speed–Accuracy Trade-Off Functions**  
**Computed Separately for the Noise-Off and Noise-On Conditions**

	Noise Off			Noise On		
	Slope	Intercept	Adjusted $R^2$	Slope	Intercept	Adjusted $R^2$
	0.017	270	93	0.01	200	87
	0.011	281	96	0.007	274	89
	0.014	192	78*	0.012	175	94
	0.021	271	95	0.008	150	83
	0.019	289	98	0.014	214	91
	0.017	200	90	0.014	192	89
	0.018	227	94	0.025	256	90
	0.007	171	97	0.012	225	96
	0.011	45	84	0.009	166	86
	0.014	214	87	0.017	247	89
	0.017	170	66*	0.019	221	80
	0.012	225	95	0.014	242	86
	0.014	342	81*	0.016	225	99
	0.008	225	83	0.009	188	68*
	0.007	42	94	0.014	200	66*
	0.008	162	94	0.007	200	87
	0.008	187	70*	0.012	225	85*
	0.010	100	97	0.013	207	90
Avg.	0.0129	200	88.4	0.0128	211	86.3
SE	0.001	80	2.25	0.001	31	1.96

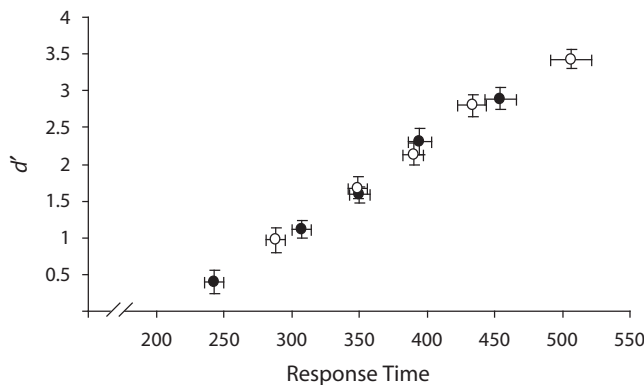
Note—All regressions were statistically significant at .05, except where indicated by an asterisk (for those regressions,  $.05 < p < .12$ ).

speed–accuracy trade-off. Although participants are faster to identify the color of a light in the presence of noise, they are also less accurate. Importantly, the slope and the intercept of the SATFs in the noise-on and noise-off conditions are practically identical. In other words, visual choice performance with and without a *facilitating* noise can be characterized by a single SATF.

These results are theoretically important because they help to distinguish between two classes of explanations that have been offered to account for the cross-modal facilitation of visual CRT: sensory and decisional. Our findings clearly favor a decisional explanation. According to this explanation, an irrelevant sound lowers the criterion that determines how much information observers use to gen-

erate the decision about the visual stimulus. By requiring less information, the responses are faster but less accurate. More importantly, the combination of benefit and cost may be predicted accurately by a single speed–accuracy function that characterizes performance with and without the irrelevant sound. If we assume that the changes in the decision criterion arise at a relatively central location, there may be nothing special about interactions between sound and light in speeded choice tasks. Other kinds of simultaneous transient stimuli—say, vibrotactile—may exert similar effects on visual CRT, also in a manner consistent with a speed–accuracy trade-off.

The present findings do not support explanations in terms of energy summation or other sensory processes



**Figure 3. Speed–accuracy trade-off functions for the noise-on (filled circles) and the noise-off (empty circles) conditions. The error bars show one standard error of the mean.**



that purport to affect visual sensitivity. Explanations in terms of sensory changes predict, contrary to the results, that an irrelevant noise should produce greater benefit than cost—that an irrelevant sound will increase speed with little or no decrease in accuracy. Were this so, then, in the presence of noise, the rate of information accumulation (slope of the SATF) should be greater, and/or the threshold (intercept) at which performance begins to exceed chance should be smaller (Lappin & Disch, 1972b). Contrary to this prediction, the SATFs in the noise-on and noise-off conditions had virtually identical slopes and intercepts.

It is interesting to note that the amount of cross-modal RT facilitation found here—40 msec—resembles the processing lag measured by simple RT in tasks of speeded auditory and visual detection (Hershenson, 1962; Woodworth & Schlosberg, 1954). This similarity might suggest the possibility that the participants responded to the noise, rather than to the color of the visual stimuli. Of course, this could not have happened on every trial or even on the majority of trials, else overall accuracy would have declined to near chance. Similar considerations render implausible the possibility that the 40 msec of facilitation in RT might result from probability summation of responses in the visual and auditory systems (sometimes proposed to account for responses to bimodal, redundantly informative stimuli; see, e.g., Colonius, 1990; Miller, 1991; Raab, 1962). This hypothesis too implies that the facilitation results from fast but random responses to the auditory stimulus on a substantial number of trials. To be sure, we cannot rule out the possibility that a few responses were made to the sound, but even if this happened, it seems unlikely that these responses substantially affected the overall pattern of results. The near identity of the SATFs obtained with and without the noise implies that similar color-based discrimination processes operated in the two conditions.

Our findings may also be relevant to interpreting the report by Stein et al. (1996) that a flash of light is judged to be brighter in the presence of irrelevant noise than in its absence. Stein et al. took this outcome as possible evidence of multisensory energy summation that increased the sensory (brightness) responses. We suggest otherwise. Although the present study measured RT but not brightness, the connection between RT and brightness is well documented. RTs in general, including CRTs, often correlate closely with measures of brightness, RT decreasing as brightness increases (e.g., Nissen, 1977; Pins & Bonnet, 1996; Schweickert, Dahn, & McGuigan, 1988). Thus, it is at least plausible to suggest that the acoustic enhancement of brightness ratings, like the acoustic facilitation of visual choice RT in selective attention, may represent a shift in decision criteria.

Odgaard et al. (2003) came to a similar conclusion about the enhancement of brightness ratings. They showed that the increase in brightness ratings in the presence of noise depends on the proportions of noise-on and noise-off trials. Specifically, the enhancement in brightness ratings disappeared when the percentage of

noise-on trials dropped from 50% to 25%. It is unlikely that an energy summation mechanism will be sensitive to changes in stimulus probabilities, whereas decision processes often are (e.g., Hansen & Well, 1984; Tanner, Haller, & Atkinson, 1967). Furthermore, Odgaard et al. (2003) found that the enhancement disappeared when the task changed from a one-interval paradigm of brightness rating (of individual lights or sound-light combinations) to a two-interval paradigm of forced choice (comparing brightness of lights alone with brightness of sound-light combinations). In many instances, two-alternative forced choice designs are less vulnerable than are single-stimulus designs to decisional shifts or response biases (Green & Swets, 1966).

Although the auditory enhancement of visual CRT and, by implication, perhaps the auditory enhancement of brightness may reflect shifts in decisional criteria, we are not suggesting that all cross-modal interactions involving intensity processing arise solely from decision processes. As has already been mentioned, a few studies have reported small sound-induced increases in visual sensitivity in tasks of absolute detection (Bolognini et al., 2005; Frassinetti et al., 2002). In tasks of bimodal detection, however, evidence of sensory enhancement is not universal (Brown & Hopkins, 1967; Loveless et al., 1970; Mulligan & Shaw, 1980).

More striking is the stronger evidence of a sensory basis for the complementary effects of irrelevant light on auditory intensity processing. For example, where sound-induced enhancement of brightness ratings depends on the probability that a sound will accompany the light and on the psychophysical method (Odgaard et al., 2003), light-induced enhancement of loudness ratings resists these manipulations: An irrelevant light increased loudness ratings even when stimulus probabilities varied and increased relative loudness comparisons even when the task was two-alternative forced choice (Odgaard, Arieh, & Marks, 2004). Furthermore, the presence of an irrelevant light seems to enhance, at least slightly, the detectability ( $d'$ ) of weak sounds (Lovell, Stein, & Wallace, 2003). Finally, there is evidence of cross-modal interactions involving other pairs of sensory modalities. In the case of auditory and tactile stimuli, the interactions may take the form of cross-modal masking: Using a psychophysical method that lacked control of criterion, Gescheider and Niblette (1967) reported that auditory clicks actually decreased sensitivity (increased thresholds) to vibratory taps, and vibratory taps decreased, albeit to a smaller extent, threshold sensitivity to auditory clicks. In a subsequent study, Gescheider, Herman, and Phillips (1970) used both criterion-controlled and criterion-uncontrolled methods to show that a substantial portion of the cross-modal masking may be attributed to shifts in criterion. Nevertheless, residual cross-modal (sensory) masking remained.

If an irrelevant light can augment loudness in tasks of intensity rating and paired comparison, we may expect an analogous enhancement to be evident in a task of speeded choice: An irrelevant light may increase the slope and/or decrease the intercept of the SATF in an auditory choice

task. In this regard, the speed–accuracy technique may be especially effective and broadly applicable as a tool for distinguishing changes in the *quality of information* or underlying perceptual representations from changes in decisional processes. Other investigators, to be sure, have pointed to the need for controlling for possible response biases in RT tasks (e.g., Spence & Driver, 1997). In this regard, the analysis of SATFs provides an especially powerful approach.

#### AUTHOR NOTE

The authors thank Sibyl N. Arbelo for help in data collection and James Hoffman and three anonymous reviewers for their cogent comments on an earlier version of the manuscript. Correspondence concerning this article should be addressed to Y. Arieh, Department of Psychology, Montclair State University, Montclair, NJ 07043 (e-mail: ariehy@mail.montclair.edu).

#### REFERENCES

- ALAIS, D., & BURR, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257-262.
- ARIEH, Y., & MARKS, L. E. (2003). Recalibrating the auditory system: A speed–accuracy analysis of intensity perception. *Journal of Experimental Psychology: Human Perception & Performance*, *29*, 523-536.
- BERNSTEIN, I. H., CLARK, M. H., & EDELSTEIN, B. A. (1969). Intermodal effects in choice reaction time. *Journal of Experimental Psychology*, *81*, 405-407.
- BERNSTEIN, I. H., & EDELSTEIN, B. A. (1971). Effect of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*, 241-247.
- BERNSTEIN, I. H., ROSE, R., & ASHE, V. M. (1970). Energy integration in intersensory facilitation. *Journal of Experimental Psychology*, *86*, 196-203.
- BOLOGNINI, N., FRASSINETTI, F., SERINO, A., & LÀDAVAS, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*, 273-282.
- BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433-436.
- BROWN, A. E., & HOPKINS, H. K. (1967). Interaction of the auditory and visual sensory modalities. *Journal of the Acoustical Society of America*, *41*, 1-6.
- CALVERT, G., SPENCE, C., & STEIN, B. E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- CARRASCO, M., MCELREE, B., DENISOVA, K., & GIORDANO, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience*, *6*, 699-700.
- COLONIUS, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, *34*, 253-275.
- DOSHER, B. A. (1979). Empirical approaches to information processing: Speed–accuracy tradeoff functions or reaction time—a reply. *Acta Psychologica*, *43*, 347-359.
- FRASSINETTI, F., BOLOGNINI, N., & LÀDAVAS, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*, 332-343.
- GESCHIEDER, G. A., HERMAN, D. D., & PHILLIPS, J. N. (1970). Criterion shifts in the measurement of tactile masking. *Perception & Psychophysics*, *8*, 433-436.
- GESCHIEDER, G. A., & NIBLETTE, R. K. (1967). Cross-modality masking for touch and hearing. *Journal of Experimental Psychology*, *74*, 313-320.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HANSEN, R. S., & WELL, A. D. (1984). The effects of stimulus sequence and probability on perceptual processing. *Perception & Psychophysics*, *35*, 137-143.
- HERSHENSON, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289-293.
- JACK, C. E., & THURLOW, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual & Motor Skills*, *37*, 967-979.
- LAPPIN, J. S., & DISCH, K. (1972a). The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology*, *92*, 419-427.
- LAPPIN, J. S., & DISCH, K. (1972b). The latency operating characteristic: II. Effects of visual stimulus intensity on choice reaction time. *Journal of Experimental Psychology*, *93*, 367-372.
- LAPPIN, J. S., & DISCH, K. (1973). Latency operating characteristic: III. Temporal uncertainty effects. *Journal of Experimental Psychology*, *98*, 279-285.
- 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.
- LOVELESS, N. E., BREBNER, J., & HAMILTON, P. (1970). Bisensory presentation of information. *Psychological Bulletin*, *73*, 161-199.
- LUCE, R. D. (1986). *Response times*. New York: Oxford University Press.
- MCELREE, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 817-835.
- MEREDITH, M. A. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, *14*, 31-40.
- MILLER, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 160-179.
- MULLIGAN, R. M., & SHAW, M. L. (1980). Multimodal signal detection: Independent decisions vs. integration. *Perception & Psychophysics*, *28*, 471-478.
- NICKERSON, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review*, *80*, 489-509.
- NISSEN, M. J. (1977). Stimulus intensity and information processing. *Perception & Psychophysics*, *22*, 338-352.
- 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.
- ODGAARD, E. C., ARIEH, Y., & MARKS, L. E. (2004). Brighter noise: Sensory enhancement of perceived loudness by concurrent visual stimulation. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 127-132.
- OLLMAN, R. T. (1977). Choice reaction time and the problem of distinguishing task effects from strategy effects. In S. Dornic (Ed.), *Attention and performance VI* (pp. 99-113). Hillsdale, NJ: Erlbaum.
- PACHELLA, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.
- PEW, R. W. (1969). The speed–accuracy operating characteristic. In W. G. Koster (Ed.), *Attention and performance II* (pp. 16-26). Amsterdam: North-Holland.
- PINS, D., & BONNET, C. (1996). On the relation between stimulus intensity and processing time: Piéron's law and choice reaction time. *Perception & Psychophysics*, *58*, 390-400.
- POSNER, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.
- POSNER, M. I., NISSEN, M. J., & KLEIN, R. M. (1976). Visual dominance: An information-processing account of its origin and significance. *Psychological Review*, *83*, 157-171.
- RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574-590.
- SCHMIDT, R. A., GIELEN, S. C. A. M., & VAN DEN HEUVEL, P. J. M. (1984). The locus of intersensory facilitation of reaction time. *Acta Psychologica*, *57*, 145-164.
- SCHWEICKERT, R., DAHN, C., & MCGUIGAN, K. (1988). Intensity and number of alternatives in hue identification: Piéron's law and choice reaction time. *Perception & Psychophysics*, *44*, 383-389.
- SIMON, J. R., & CRAFT, J. L. (1970). Effects of irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology*, *86*, 272-274.
- SPENCE, C., & DRIVER, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*, 389-403.

- STEIN, B. E., LONDON, N., WILKINSON, L. K., & PRICE, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, **8**, 497-506.
- STEIN, B. E., & MEREDITH, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- SWENSSON, R. G. (1972). Trade-off bias and efficiency effects in serial choice reactions. *Journal of Experimental Psychology*, **95**, 397-407.
- TANNER, T. A., JR., HALLER, R. W., & ATKINSON, R. C. (1967). Signal recognition as influenced by presentation schedules. *Perception & Psychophysics*, **2**, 349-358.
- TODD, J. W. (1912). Reactions to multiple stimuli. *Archives of Psychology*, **21**(Whole No. 25), 1-65.
- VROOMEN, J., & DE GELDER, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1583-1590.
- WICKELGREN, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, **41**, 67-85.
- WOODWORTH, R., & SCHLOSBERG, H. (1954). *Experimental psychology*. London: Methuen.

(Manuscript received December 11, 2006;  
revision accepted for publication September 19, 2007.)