

Temporal integration in vision: Masking can aid detection

Dan J. Swift

Published online: 10 January 2013
© Psychonomic Society, Inc. 2013

Abstract Temporal form-part integration is the process whereby two discrete sets of stimuli, presented at different times, are perceived by the visual system as a single integrated percept. Dixon and Di Lollo (Cognitive Psychology 26(1):33–63, 1994) proposed a temporal correlation model that was able to quantitatively account for a number of findings related to both form-part integration tasks and partial report. The present study demonstrates a novel approach to form-part integration—the addition of a whole-field mask stimulus at the termination of the second set of stimuli. According to an extension of the correlation model, the mask stimulus should reduce the visual system’s response to the trailing display, thus increasing the relative overlap of the two displays, fostering integration. Experiment 1 supported this hypothesis, showing a maximum benefit when the mask followed the trailing display immediately, as opposed to after a delay of 60 ms. Experiment 2 showed that this same mask actually did yield worse performance when presented immediately after a single set of stimuli. The third experiment collected detailed data for a few observers over a larger number of mask delays. Taken together, these experiments provide a rare example of masking degrading a target stimulus, and yet aiding perception.

Keywords Temporal processing · Visual perception · Masking · Temporal impulse response function (TIRF)

Temporal integration in vision was introduced as a task to determine the time course of visible persistence. The basic task, known as *form-part integration*, is to divide a stimulus in two, presenting each part at different time intervals. The stimulus may be two parts of a nonsense syllable, consisting of small dots (Eriksen & Collins, 1967), or two parts of a matrix of dots, with one position missing (Hogben & Di Lollo, 1974). In either case, the task is to report on properties of the

combined stimulus. The general finding is that temporal integration occurs even with a temporal gap between the two sets of stimuli; the larger the gap, the less likely that integration will occur. The size of the gap has been taken as an indication of visible persistence; that is, even though the two stimuli do not physically overlap, they do overlap physiologically, because of persistence of the response to the stimuli in the visual system. Typically, integration still occurs even when the gap is in excess of 100 ms (Eriksen & Collins, 1967), suggesting that persistence must last for at least 100 ms. However, it quickly became apparent that visible persistence alone might not be sufficient to account for all of the results, such as when the two sets are presented at different luminance levels. Eriksen and Collins suggested that additional processes, such as discontinuity detectors, or concepts, such as the perceptual moment, may need to be invoked in order to account for their findings.

Dixon and Di Lollo (1994) proposed a theory of temporal integration that seems to account nicely for a range of findings, from both form-part integration and partial-report tasks. According to their model, when two stimuli or sets of stimuli are presented in quick succession, the visual system analyzes the correlation between the responses to the two sets of stimuli. If the correlation is high enough, then integration occurs; otherwise, the two stimuli are segregated, and performance on the experimental task becomes much more difficult. This is similar to saying that the relative overlap of the visual-system responses to the two sets of stimuli is what promotes integration, rather than the absolute amount of overlap. Two sets of stimuli that start at the same time but end at different times may not integrate, due to the amount of nonoverlap. A key component in this model is a temporal impulse response function (TIRF) that determines the level of the visual-system activity in response to the stimulus at any given time. Presumably, the time constant of the TIRF relates to the duration of visible persistence, although that point is perhaps arguable. Recent visual evoked potential data (Akyürek, Schubö, & Hommel, 2010) have suggested that temporal integration is a very early, preattentive process, which lends some support to the idea that the time constant of the TIRF relates to visual

D. J. Swift (✉)
Behavioral Sciences Department, University of
Michigan–Dearborn, Dearborn, MI 48128, USA
e-mail: dswift@umich.edu

persistence. The main point, however, is that, without a comprehensive model of temporal integration, we can make inferences neither about visible persistence nor about how the visual system integrates disparate stimuli.

The primary goals of the present study were to (1) test the viability of the Dixon and Di Lollo (1994) correlation model of temporal integration through use of a masking paradigm, and (2) document a paradigm in which a mask with a short delay improves overall performance while still degrading perception of the target stimuli. The concept is this: According to the correlation model, failure to integrate two stimuli is often not due to the lack of overlap between the visual system's responses to the two sets of stimuli; rather, the amount of nonoverlap between the two sets may outweigh the overlap. The novel idea in the present study is to introduce a third stimulus—a whole-field mask. Since the noise nature of the mask stimulus would seem to preclude constructive integration, regardless of its temporal correlation with the preceding stimuli, its main effect would be to decrease the trailing, nonoverlapping response to the second set of stimuli, hence increasing the temporal correlation value of the two stimulus sets. The mask should, therefore, improve performance on the temporal integration task, and the earlier that the mask is presented, the greater that improvement will be.

This study comprises three separate experiments. The first demonstrates the basic effect of masking improving performance, using a sample of naïve observers, thus permitting the use of inferential statistics to extend the findings. The second tests whether the mask stimulus is actually functioning as a mask, by using it in a context with just one set of stimuli. The third uses trained observers to extend the findings to multiple mask delays, in order to determine the delay at which the mask no longer has an effect.

Since pilot data indicated a small effect that would benefit from standard tests of statistical significance, data were collected from a large group of naïve observers with no prior experience with visual psychophysical tasks. This was feasible because, when temporal integration clearly occurs, the missing matrix position is easy to detect, and the task therefore requires little training.

Experiment 1

Method

Participants The participants in this study consisted of 25 students in introductory psychology, most of them of a traditional age for undergraduates, who were fulfilling a course requirement for experimental participation.

Stimuli All of the stimuli were presented on a 17-in. or a 19-in. CRT monitor, set at 800×600 resolution and a 100-Hz

frame rate. The stimuli consisted of individual 1.5-deg solid white squares with a luminance approximately equal to 100 cd/m^2 , on a gray whole-screen background with a luminance approximately equal to 50 cd/m^2 . The center-to-center spacing was 2.25 deg, and the viewing distance was approximately 75 cm. On any given trial, seven or eight stimuli were presented for 100 ms, followed by a second set of seven or eight stimuli, also presented for 100 ms, such that the 15 stimuli filled all but one position of a combined 4×4 matrix. The stimulus matrix was perfectly rectangular, for otherwise the concept of the missing position would have been less well defined. The stimulus onset asynchrony (SOA) between the two sets varied from 50 to 120 ms. A third stimulus, the whole-field mask, was presented either immediately after offset of the second set of squares or with a delay of 60 ms, for a duration of 1 s, which was typically less than the response time. The mask consisted of a field of achromatic squares of random luminance from 0 to 100 cd/m^2 . Each square was 0.164 deg per side, and the entire field was centered on the screen and extended 12.4 deg wide by 13.1 deg high. The entire stimulus sequence is illustrated in Fig. 1.

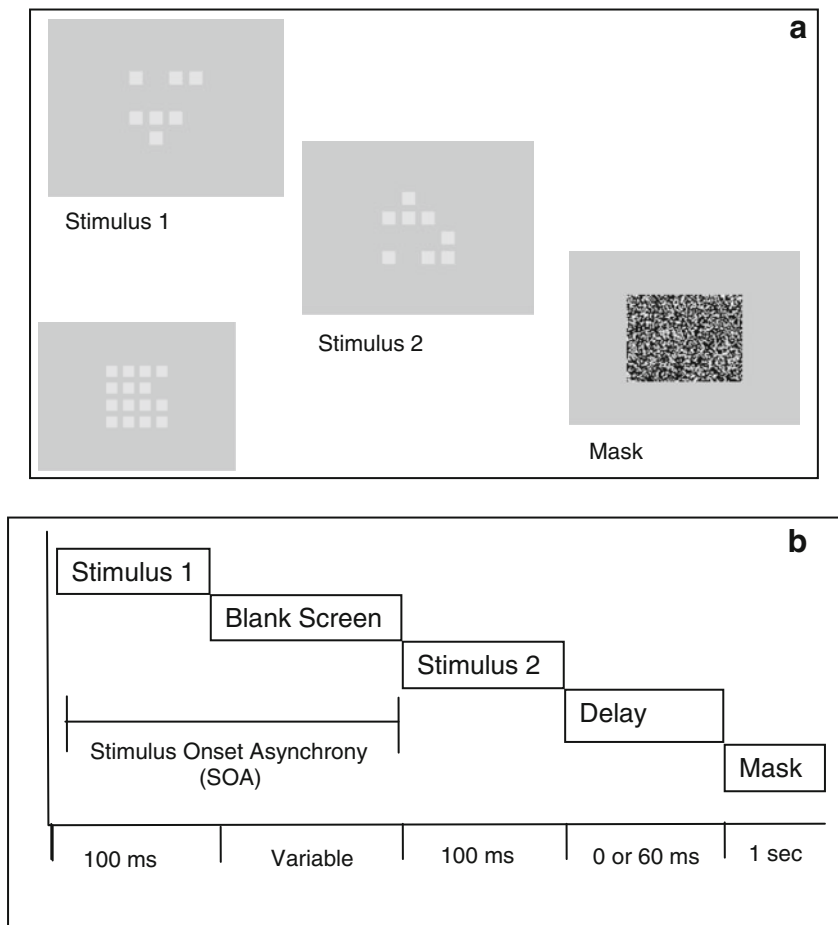
Design A total of seven SOAs, from 50 to 120 ms, were combined with two mask delays, 0 and 60 ms, resulting in a two-way (7×2) repeated measures design.

Procedure The experiment was controlled by the Vision-Works software (Swift, Panish, & Hippensteel, 1997). Participants were instructed to press the key corresponding to the matrix position that did not receive a square. The keyboard was set up with blue tape on the keys “3”–“6,” “E”–“Y,” “D”–“H,” and “C”–“N” on a QWERTY keyboard, such that there was a rough mapping of 16 keys to the 16 possible matrix positions; the positions of the keys, of course, were not perfectly rectangular. A set consisted of 140 trials ($7 \text{ SOAs} \times 2 \text{ delays} \times 10 \text{ replications}$), interleaved randomly. Pacing was automatic, with only about 2 s between trials. Participants were informed that they should do as many sets as possible in a 2-h period and that the first set would be treated as practice. Rests were permitted after each set, and a longer, 15-min rest was provided halfway through. Most of the participants completed six or seven sets in addition to the practice set.

Results

Of the 25 participants, only two produced results that were deemed unusable, as a result of an a priori decision to establish a cutoff of at least 50 % correct on each of the two easiest conditions (50-ms SOA). In fact, the two participants who did not meet this criterion were well below it,

Fig. 1 (a) A stimulus sequence is shown from top left to bottom right. The insert at the lower left is the integration of the first two sets of stimuli. The correct response for this trial would be the “Y” key, indicating the rightmost position of the second row. (b) Timing is automatic; after the observer makes a response, the fixation point appears for 2 s, followed by the stimuli as indicated above. The blank screen is variable, and may be absent, as when Stimulus 2 directly follows Stimulus 1, or even overlaps in time with Stimulus 1



with one at 25 % and the other at 3 % correct. Perceptual learning is always an issue, especially when dealing with naïve observers, since such learning may confound the effects of the independent variables. However, since all conditions were interleaved, any learning that did take place would have affected all conditions equally. An analysis of the data indicated that after the initial practice round, the overall improvement was minimal, and that after the first three rounds, there was no more improvement. An analysis based solely on the final three rounds showed slightly higher numbers for percent correct across the board, but a virtually identical pattern of results, so only the data on the full six to seven rounds after the initial practice round from the 23 remaining participants are analyzed here and shown in Fig. 2.

As can be readily seen, both main effects were significant. As SOA increased, performance decreased: $F(6, 126) = 291.01, p < .001, \eta^2 = .933$. Also, the absence of a delay between the offset of the second set of stimuli and the onset of the mask yielded better performance than did delaying the mask by 60 ms: $F(1, 21) = 97.828, p < .001, \eta^2 = .823$. While the effects were similar for the two delay conditions, the interaction between delay and SOA was also significant: $F(6, 126) = 7.251, p < .001, \eta^2 = .257$. This appears to reflect

the larger advantage for the 0-delay mask at the intermediate SOA values.

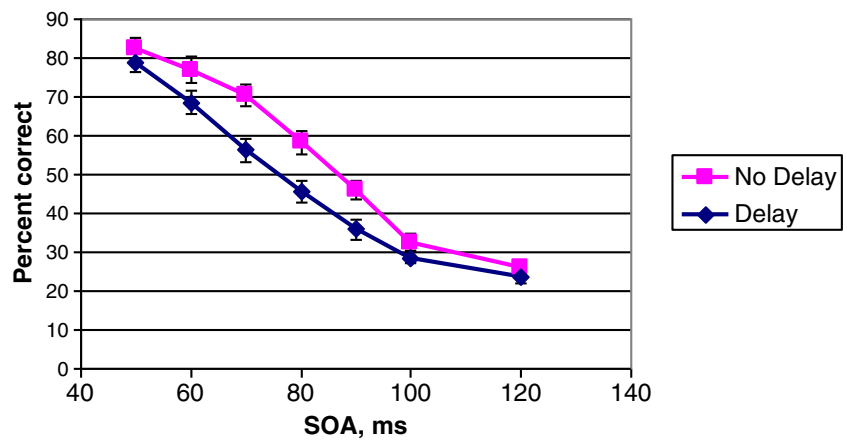
It should be noted that only the 100-ms and 120-ms SOA conditions resulted in no physical overlap of the stimuli. Therefore, it could be argued that these conditions provide a stronger case for the necessity of some form of visible persistence or visual memory. Hence, an analysis of variance (ANOVA) was performed on these SOAs alone. In this case, the effect of delay—while reduced, as can be seen in Fig. 2—was still highly significant, $F(1, 22) = 34.4, p < .001$.

These data were then fit to the Dixon and Di Lollo temporal correlation model. The model includes several parameters, all of which relate directly to various components of the model. The model starts with the assumption of a TIRF in the form of a gamma function of time (t):

$$g(t) = \left(\frac{t}{\tau}\right)^{n-1} \frac{e^{-t/\tau}}{\tau(n-1)!}$$

The exponent (n) was fixed at 10, following Dixon and Di Lollo (1994). The time constant (τ) was one of the parameters to be fit. This TIRF is convolved with the stimulus to produce the visual system’s response. The next step

Fig. 2 Percent correct as a function of stimulus onset asynchrony (SOA). Error bars show ± 1 standard error



in the model is to correlate the visual-system responses of the two sets of stimuli. The correlation is kept as a running total, with more recent values weighted more heavily, according to the decay function: $e^{-(t-t_i)/a}$. The decay parameter (a) was fixed at 340 ms, similar to that used by Dixon and Di Lollo. The next step is to compute the probability of integration, based on a comparison of the correlational value with a criterial correlation (R_c). A normal distribution of correlation values is assumed, with variance $1/(ns - 3)$, in which ns is the effective sample size. Finally, a decision rule is needed:

$$P_c = \frac{1}{16 - 15d} P_i + \frac{1}{16 - 7.5d} (1 - P_i)$$

in which P_c is the probability of a correct response, and P_i is the probability of integration. The d parameter indicates the probability of detecting each individual dot in the trailing display, which is typically close to 1. The value 16 refers to the total number of element positions, 15 is the number of actual elements, and 7.5 is the average number of elements in the trailing display. Note that this formula assumes that it is unlikely that the observer will respond with an element that was present in the trailing display. Thus, if P_i equals 0 and $d = 1$, the probability of a correct response equals $1/8.5$ (12 %), not $1/16$, which does help account for the fact that even at long SOAs, the present results show percent-correct values above 20. The remaining difference between the 12 %- and 20 %-correct values presumably arises from some memory of the locations of the stimuli in the leading display and/or from actual temporal integration due to the long tail of the TIRF.

Of the parameters, two were treated as free parameters—the time constant (τ) of the TIRF and the criterial coefficient for temporal integration (R_c). The best fit was accomplished with Microsoft Excel, using a generalized reduced gradient model (Lasdon, Waren, Jain, & Ratner, 1978). The fit was confirmed by a program written by the author, using a simpler, iterative algorithm. Fits were performed for the 60-ms-delay

condition, in which the stimuli were separated by a larger time interval from the mask and were less affected by it. The parameters were then applied to the 0-ms-delay condition with no further parameter optimization. The details of the Dixon and Di Lollo (1994) model can be found in their original article; the extensions of the model used to account for masking are explained here.

It is assumed that the visual-system response to the mask stimulus, like the response to every other stimulus, is mediated by a TIRF (Watson, 1986), so that TIRFs with identical form and time constants were applied to both sets of stimuli and to the mask. It was further assumed that the visual-system response to the stimulus would be reduced by the mask. The precise mathematical form of this reduction may be in question, but it appears that as long as white noise is used (i.e., the noise power is spread out over the frequency domain for all orientations), additive masking may be assumed (Majaj, Pelli, Kurshan, & Palomares, 2002; Pelli & Farell, 1999; Swift & Smith, 1983). Therefore, at any given time, the response to the mask was subtracted from the response to the stimulus, with a minimum of 0 stimulus response. This fits with the idea of object substitution (Enns, 2004), though in most studies of masking the task is detection. In this case, the idea is that the suprathreshold response to a stimulus is diminished until the stimulus is phenomenologically replaced by the noise mask. Then, following the Dixon and Di Lollo model, the temporal correlation, probability of integration, and hypothesized percent correct were computed, on the basis of the visual-system response to the stimuli after the masking effect was computed. One additional difference between the Dixon and Di Lollo model and the present computation is that in the original model, a moving Gaussian window was used to weight recent responses more heavily than earlier ones. The same window was applied here, with a constant width of 340 ms, but only after a z-score transformation of the responses to ensure equal weighting prior to the application of the window. This turned out to have only a small quantitative effect, and no real qualitative one.

For comparison purposes, the computation was done with a slightly different computational model of masking. Instead of simply subtracting the response to the mask stimulus from the response to the target stimulus, the response to the mask stimulus was multiplied by the response to the target and then subtracted; hence, the amount of masking was proportional to the current level of target response.

The fits to the data are shown in Fig. 3, for the same gamma function TIRF used by Dixon and Di Lollo.

It can readily be seen that both types of masking are qualitatively in line with the data, with a slightly better fit for the multiplicative mask. However, this may be misleading. In order to minimize the number of free parameters, no additional parameter was used to adjust the amount of masking. In essence, this resulted in a masking factor of 1, meaning that the assumption was made that the visual system’s response to the mask was equal to the response to the target stimulus. Good quantitative fits can be achieved with either masking type if this assumption is relaxed and the masking factor is adjusted. The model fits are also consistent with the interaction, which showed the largest difference between the 0- and 60-ms delays at intermediate SOA levels.

A further analysis was done with a different TIRF—a biphasic function based on Bergen and Wilson (1985), as is

illustrated in Fig. 4. The parameter values for both TIRFs, along with descriptions, are provided in Table 1. There are only minor differences in the parameter values for the two different masking functions.

It is apparent that both TIRFs give reasonable fits with both additive and multiplicative masking. However, the TIRFs are different functions; a direct comparison of the two TIRFs is shown in Fig. 5, using the respective fit parameters shown in Table 1 for additive masking. The width of the gamma function at half amplitude is approximately 200 ms. The half-amplitude width of the biphasic function is about 100 ms for the positive portion, and about 150 ms for the negative portion. The width of the gamma-function TIRF is larger than that found by Dixon and Di Lollo (1994), and even more so than that found by Watson (1986). Two points are relevant—both of which were made by Dixon and Di Lollo. First, with regard to the narrower TIRF found by Watson (1986), their claim was that the TIRF may well widen at higher stages of processing, and this task likely represents a higher stage. Second, they acknowledged that “a variety of different parameter values would provide fits of similar quality” (p. 60). This is equivalent to saying that the parameters interact and that the confidence interval for each parameter is fairly wide. The present analysis confirmed this finding.

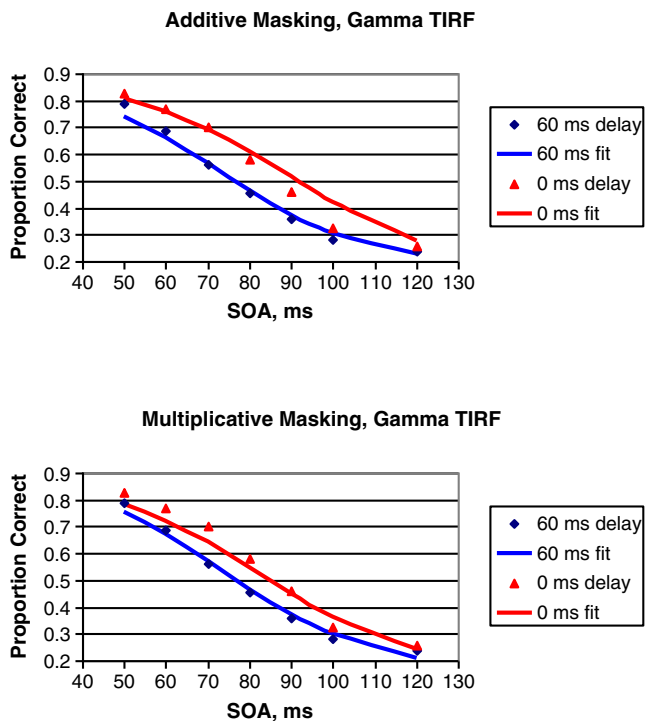


Fig. 3 Model fit for a gamma temporal impulse response function (TIRF; Dixon & Di Lollo, 1994) for the 60-ms-delay condition; no further fitting was done for the 0-ms-delay condition. Two masking functions are shown: additive (top) and multiplicative (bottom; see the text for details)

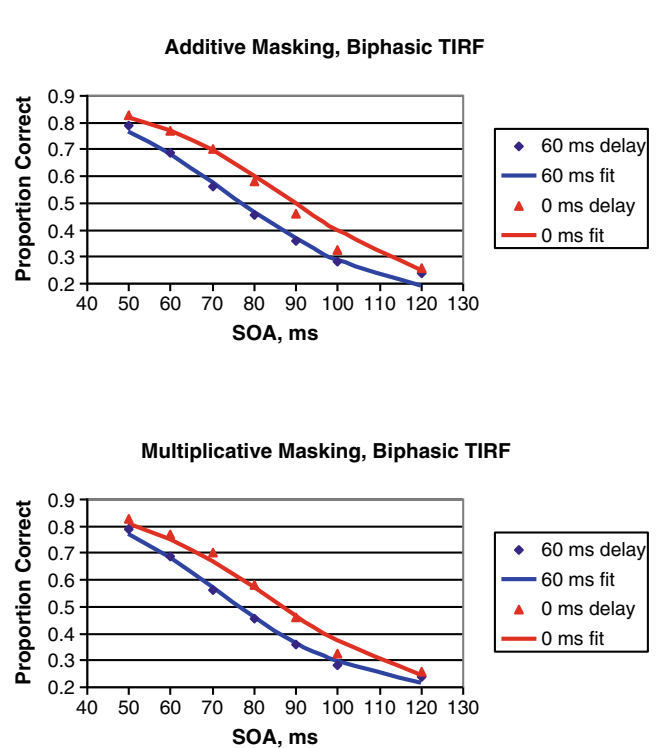


Fig. 4 Model fit for a biphasic TIRF (Bergen & Wilson, 1985) for the 60-ms-delay condition; no further fitting was done for the 0-ms-delay condition. Two masking functions are shown: additive (top) and multiplicative (bottom; see the text for details)

Table 1 Parameter fits based on a gamma temporal impulse response function (TIRF) and a biphasic TIRF for both additive (+) and multiplicative (\times) masking, all based on the Dixon and Di Lollo (1994) correlation model of temporal integration

Parameter	Status	Description	Monophasic TIRF Values		Biphasic TIRF Values	
			+	\times	+	\times
τ	Fit	Time constant of gamma function (ms)	31.14	25.78	29.25	34.17
Exponent	Constant	Exponent; for monophasic (gamma) TIRF, larger values indicate smaller amounts of positive skew	10	10	6	6
W_τ	Constant	Time constant of sliding temporal window (ms)	340	340	340	340
ns	Constant	Number of samples, used to compute variance	5	5	5	5
Rc	Fit	Minimum correlation coefficient required for temporal integration	.64	.54	.42	.42
d	Constant	Dot detection rate (probability that any given element of the matrix will be detected)	.99	.99	.99	.99

Discussion

It is, of course, not surprising that good fits were obtained in the 60-ms-mask delay conditions—any reasonable model should be able to fit seven data points with two free parameters. What is interesting is that once the data for the 60-ms-mask delay were fit, the fit for the 0-ms-delay condition was also quite good, with no additional free parameters. It is also interesting to observe that while both TIRFs give equivalent fits to the data and yield roughly equivalent time constants, the width of the positive lobe of the biphasic function is considerably narrower than the gamma function. It does not appear that temporal integration, at least in the present context, can be used to identify the form of the TIRF. However, identifying the form of the TIRF was not the primary goal of this study, but rather to further test the Dixon and Di Lollo correlation model, and this model seems to give a very good prediction, regardless of the form of the TIRF.

One of the strengths of the present test of the correlation model is that it makes a nonintuitive prediction—that is, that masking can aid perception. Intuitions, however, are not always the best guide. In fact, there are other reported instances of masking aiding perception, but these tend to be in fairly specialized circumstances, often when the mask stimulus combines with the target stimulus in a constructive way. One such situation is near-threshold pedestal detection (Nachmias & Sansbury, 1974; Swift & Smith, 1984), and another involves vernier offset discrimination (Hermens, Herzog, & Francis, 2009). However, in the present instance, integration is facilitated and performance improved precisely *because* masking is effective for reducing the visual-system response to the second set of stimuli. This mask was random on a small scale, but homogeneous on a large scale, in order to equalize its effect at all matrix positions, thus avoiding spurious correlations that might improve performance. Herzog, Fahle, and Koch (2001) found that a homogeneous mask was beneficial to performance in their shine-through illusion, but reducing the homogeneity caused performance to deteriorate.

Examples of perception being improved by reducing the visual system's response to the stimulus that is actively being detected are at best rarer in the literature, but that is exactly what the present implementation of the Dixon and Di Lollo model predicts, and what the present data show. The closest example of this is probably an experiment by Herzog, Parish, Koch, and Fahle (2003) in which a mask appeared to affect the second of two sequentially presented verniers, thus biasing performance in favor of the first vernier. However, their study showed an overall improvement in performance only if the task was defined by the first vernier. In a sense, the mask stimulus was masking a mask (the second vernier); in the present case, the mask is masking the second stimulus, which is a target stimulus. Since the mask stimulus here was presented at the termination of the trailing set of stimuli, it had an SOA of 100 ms. Previous research has shown that noise masking has a greatly reduced effect by 100 ms (Enns, 2004), which reinforces the finding that the mask that has the greatest effect on the trailing set of stimuli actually has the largest beneficial effect on the form-part integration task.

An issue arises with the use of the term *masking*. Since the only predicted and observed effect of the visual noise was to facilitate performance, a case can be made that this “mask”

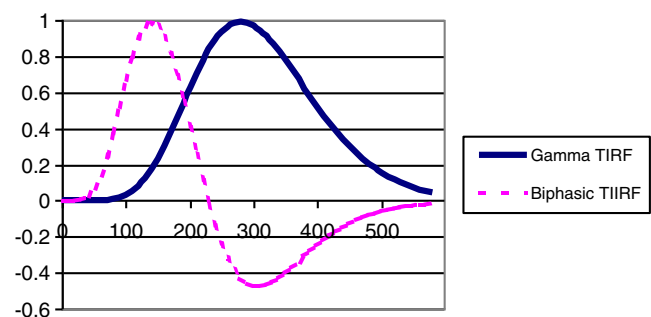


Fig. 5 The two TIRFs used in the model fits. The values for each TIRF are normalized to a positive peak of 1. The x-axis represents time in milliseconds

stimulus is simply misnamed—that is, that it in fact is not masking anything. The model, of course, makes its prediction on the basis of the visual noise masking the second set of target stimuli, but there is no independent evidence of this. Therefore, a second experiment was devised to test the masking effect directly. In order to make it as similar as possible to Experiment 1, the task from the observer's point of view was identical—to detect the position of the missing element. This task was not a perfect analogue of the task in Experiment 1, since the task has been made easier by virtue of presenting a stimulus set in which the stimuli are physically integrated, as opposed to requiring the integration to take place in the visual system. In order to make the task more difficult, it was necessary to present the stimuli for a much shorter duration than the 100 ms used in Experiment 1. Alternative controls were considered and rejected. Any task that included just seven or eight squares presented for 100 ms (i.e., identical to Stimulus 2 from Exp. 1) would suffer from requiring different instructions for the observer, and thus not being identical to Experiment 1 in that regard. Furthermore, the stimuli were superthreshold, so that if, for example, observers were asked to indicate whether a given position had contained a stimulus, the task would suffer from ceiling effects. The original form-part integration task could have been made more difficult by reducing the contrast of the stimuli, but contrast reduction has been shown to increase the time constant (Di Lollo & Bischof, 1995), and thus contrast reduction would be a confounding factor. Therefore, the control with a shorter stimulus presentation was deemed the closest possible equivalent to Experiment 1 for purposes of measuring the masking effect.

Experiment 2

Method

The method was identical to that used in Experiment 1, except that the stimuli consisted of a single set of 15 elements in an array of 16 possible positions. The presentation time of the single set was only 20 ms, another difference from Experiment 1, but one that was required to avoid a complete ceiling effect. The independent variable in this case was mask delay—0, 30, 60, 120, or 900 ms. As before, all conditions were run in blocks of ten replications, with all conditions interleaved. The participants were 16 different students in introductory psychology, who were also fulfilling a course requirement for experiment participation.

Results and discussion

The results are shown in Fig. 6 for all delay conditions.

It is clear that a significant masking effect occurs when the mask immediately follows the target stimulus. Masking

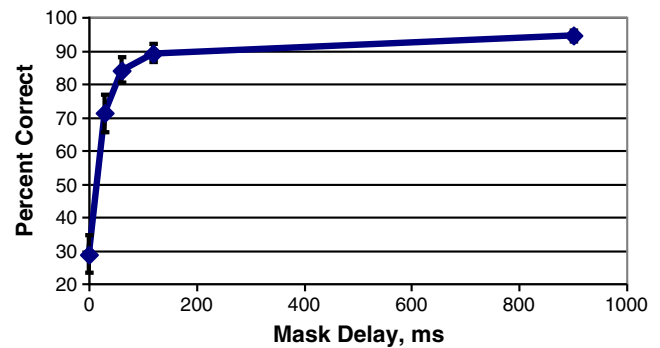


Fig. 6 Effect of mask delay on percent correct. Error bars show ± 1 standard error

is greatly reduced with a 30-ms delay, and is almost eliminated by 60 to 120 ms. These results are roughly in line with previous results on masking. Enns (2004), using a partial-report paradigm with a noise mask, showed that masking is not effective at SOAs beyond about 50 ms. Thus, while the stimulus and task conditions could not be quite identical to those in Experiment 1, the results here do support the main argument that the results in Experiment 1 can be explained by the masking effect of the noise stimulus.

One weakness of Experiment 1 was that only two SOAs were used. The reason for this was that the primary goal was to demonstrate a benefit of masking under the best circumstances with a large group of naïve observers who were available for only a limited time. Nonetheless, it would be useful to determine the effect of varying the timing of the onset of the mask stimulus, in part to see whether it would be consistent with the results of Experiment 2.

Experiment 3

Method

The method was identical to that used in Experiment 1, with the only difference being that additional conditions were tested: all the mask delays used in Experiment 2. The observers included the author and two student volunteers, one with little experience on this task, and one with extensive experience. As before, all conditions were run in blocks of ten replications, with all conditions interleaved.

Results and discussion

The results are shown in Fig. 7 for all delay conditions, including the no-mask condition. Graphs are not shown for the individual observers, since they were all quite similar. The only difference was slightly lower percent-correct values across all conditions for the least experienced observer. Masking presented immediately upon termination of the second set

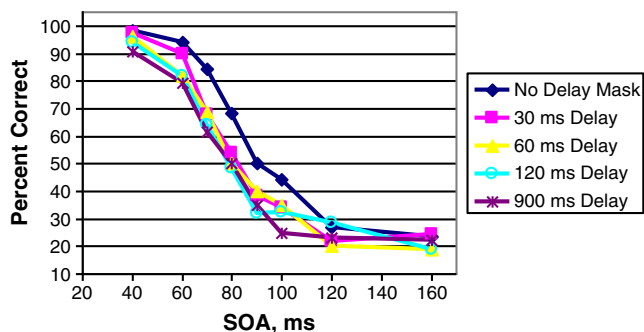


Fig. 7 Data for three observers, including the author, over a wider range of masking delays and SOAs

of stimuli was more effective than any delay, including one as long as 900 ms, which is essentially no mask at all, since almost all responses were made while the mask was still present.

Despite including only three observers, the data were analyzed via ANOVA. Both SOA [$F(7, 14) = 60.74, p < .001, \eta^2 = .968$] and mask delay [$F(4, 8) = 13.12, p = .001, \eta^2 = .868$] were highly significant, but the interaction was not significant. Both variables also showed a significant linear trend [SOA, $F(1, 2) = 324.80, p = .003, \eta^2 = .994$; delay, $F(1, 2) = 57.44, p < .014, \eta^2 = .966$]. The effect of delay, averaged across all SOA conditions, is shown in Fig. 8. It mirrors the results of the control experiment (Exp. 2), in that the largest effect is seen with no delay of the mask, with a large decrease in effect as mask delay increases. Of course, this result is opposite in the sense that, in the control experiment, percent correct increased with increasing mask delay, whereas here, percent correct decreases with increasing mask delay, but this masking facilitation effect is just what the model predicts. Figure 9 shows the quantitative relationship between degree of masking and degree of facilitation. The relationship is nearly perfect, which reinforces the appropriateness of the control and the overall

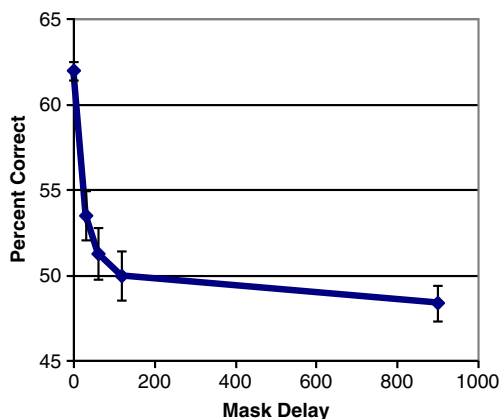


Fig. 8 Percent correct as a function of mask delay, averaged across three observers for all SOAs. Error bars show ± 1 standard error, adjusted for the observer effect

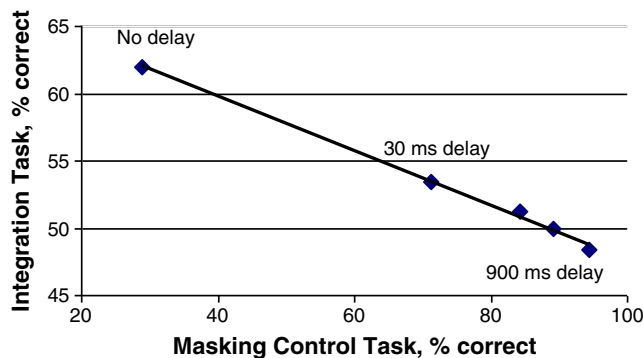


Fig. 9 Relationship between degree of masking on the control task and degree of facilitation resulting from the mask stimulus on the temporal integration task. The best-fit line is shown, with $r = -.998$

interpretation that the facilitation effect results from masking of the trailing portion of the target stimulus.

General discussion

The results suggest that the role of the mask stimulus is to reduce the visual system’s response to the trailing stimulus, thereby promoting integration. It has been suggested that masking may not operate in this way. Rather, it may operate at a higher level, disrupting the feedback (reentrant) connections to the lower level. In effect, the visual system’s low-level response to the target stimulus remains unaffected by the mask, but at some point the conscious percept of the target stimulus is replaced by that of the mask (Fahrenfort, Scholte, & Lamme, 2007). This is somewhat similar to Sperling’s (1971) suggestion that the effect of a second stimulus in a sequence is to terminate the visibility of the first stimulus and replace it with the second. One difference is that Sperling is less specific about a possible transition phase, whereby the visibility of the target stimulus is reduced by the mask.

There is no logical way to rule out this “reentrant” view of masking, but it can be tested empirically, at least to the degree that it can account for the present data. According to the reentrant view, there should be no reduction in visual-system response at all; rather, at some point the response to the mask replaces the response to the target as a conscious percept. The best way of modeling the reentrant idea is to use Sperling’s (1971) replacement idea without any transition. At some point, the response to the target stimulus is simply replaced by the response to the noise stimulus, presumably when the visual system’s response to the noise stimulus exceeds that to the target. At this point, the response of the target stimulus is assigned a value of 0. This analysis was performed accordingly. As can be seen in Fig. 10, this model of masking gives the opposite prediction from the present finding that masking facilitates integration.

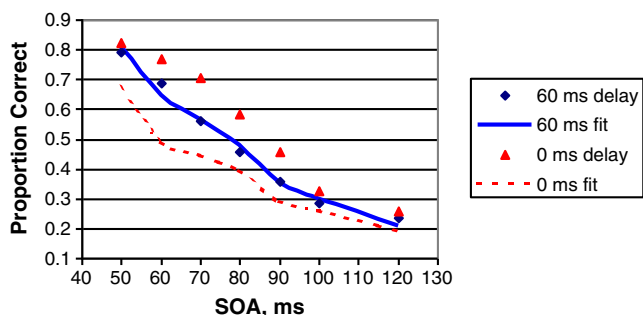


Fig. 10 An attempt to fit the data with an object substitution model of masking, with no transition of reduced target response

Conclusion

This study may be the first demonstration of using a third stimulus to reduce the persistence of a second (target) stimulus, thereby improving performance on a temporal integration task. It should be noted that the present study is not the first time that masking has been examined with respect to a temporal integration task. Groner, Groner, Bischof, and Di Lollo (1990) used a frame around the second set of stimuli, which had the effect of reducing the persistence of the first set of stimuli, thereby reducing overall performance. In the present study, on the other hand, a mask was used to improve task performance by degrading stimulus persistence.

The present use of a novel masking paradigm provides support for the Dixon and Di Lollo correlation model of temporal integration. This, of course, does not preclude other explanations of temporal integration. For example, as Dixon and Di Lollo (1994) acknowledged, the model proposed by Groner, Bischof, and Di Lollo (1988) is mathematically similar to the temporal correlation model, although the underlying assumptions about the interval process are different. Groner et al. (1988) proposed that integration was related to the amount of overlap in the visual system's response to the two stimuli, divided by the total amount of energy in the two sets of stimuli. It appears that this model would also be supported by the present data, from the assumption that the mask stimulus reduces the energy in the trailing stimulus, thus making the relative overlap greater.

Another model that has been used to account for temporal integration is a neural-network model known as the *boundary contour system* (Francis, 1996). That model postulates that resonance is responsible for visible persistence, but that various lateral inhibition processes account for such effects as reduced integration with increasing duration of the leading and trailing displays. It is difficult to predict how the boundary contour system would account for the present data, as “the simulations predict general trends; they do not predict the display parameters necessary to produce these results” (Francis, 1996, p. 1209). The model incorporates inhibition of reverberating circuits as part of

an account for masking, so it is certainly possible that the present data could be accounted for by this model.

With regard to the main goals of the study, the following conclusions emerge: (1) The Dixon and Di Lollo temporal correlation model is supported by the present findings; (2) although a precise temporal time constant may be difficult to determine with certainty, due to the presence of other parameters in the model, the data do suggest a time constant of 30 ms for the gamma TIRF, resulting in a width of about 200 ms at half amplitude; and (3) a novel paradigm has been demonstrated in which masking improves performance.

Author note The author wishes to acknowledge the constructive criticism of four reviewers and the editor. Also, completion of this article was supported by a sabbatical from the University of Michigan–Dearborn.

References

- Akyürek, E. G., Schubö, A., & Hommel, B. (2010). Fast temporal event integration in the visual domain demonstrated by event-related potentials. *Psychophysiology*, *47*, 512–522. doi:10.1111/j.1469-8986.2010.00962.x
- Bergen, J. R., & Wilson, H. R. (1985). Prediction of flicker sensitivities from temporal three-pulse data. *Vision Research*, *25*, 577–582. doi:10.1016/0042-6989(85)90163-4
- Di Lollo, V., & Bischof, W. F. (1995). Inverse-intensity effect in duration of visible persistence. *Psychological Bulletin*, *118*, 223–237. doi:10.1037/0033-2909.118.2.223
- Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology*, *26*, 33–63. doi:10.1006/cogp.1994.1002
- Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, *44*, 1321–1331. doi:10.1016/j.visres.2003.10.024
- Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, *74*, 476–484. doi:10.1037/h0024765
- Fahrenfort, J. J., Scholte, H. S., & Lamme, V. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*, 1488–1497. doi:10.1162/jocn.2007.19.9.1488
- Francis, G. (1996). Cortical dynamics of visual persistence and temporal integration. *Perception & Psychophysics*, *58*, 1203–1212. doi:10.3758/BF03207553
- Groner, M. T., Bischof, W. F., & Di Lollo, V. (1988). A model of visible persistence and temporal integration. *Spatial Vision*, *3*, 293–304. doi:10.1163/156856888X00177
- Groner, R., Groner, M. T., Bischof, W. F., & Di Lollo, V. (1990). On the relation between metacontrast masking and suppression of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 381–390. doi:10.1037/0096-1523.16.2.381
- Hermens, F., Herzog, M. H., & Francis, G. (2009). Combining simultaneous with temporal masking. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 977–988. doi:10.1037/a0014252
- Herzog, M. H., Fahle, M., & Koch, C. (2001). Spatial aspects of object formation revealed by a new illusion, shine-through. *Vision Research*, *41*, 2325–2335. doi:10.1016/S0042-6989(01)00122-5

- Herzog, M. H., Parish, L., Koch, C., & Fahle, M. (2003). Fusion of competing features is not serial. *Vision Research*, *43*, 1951–1960. doi:10.1016/S0042-6989(03)00278-5
- Hogben, J. H., & Di Lollo, V. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. *Vision Research*, *14*, 1059–1069. doi:10.1016/0042-6989(74)90202-8
- Lasdon, L. S., Waren, A. D., Jain, A., & Ratner, M. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Transactions on Mathematical Software*, *4*, 34–50. doi:10.1145/355769.355773
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*, 1165–1184. doi:10.1016/S0042-6989(02)00045-7
- Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research*, *14*, 1039–1042. doi:10.1016/0042-6989(74)90175-8
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America. A*, *16*, 647–653. doi:10.1364/JOSAA.16.000647
- Sperling, G. (1971). Information retrieval from two rapidly consecutive stimuli: A new analysis. *Perception & Psychophysics*, *9*, 89–91. doi:10.3758/BF03213034
- Swift, D., Panish, S., & Hippensteel, B. (1997). The use of *Vision-Works™* in visual psychophysics research. *Spatial Vision*, *10*, 471–477. doi:10.1163/156856897X00401
- Swift, D. J., & Smith, R. A. (1983). Spatial frequency masking and Weber's Law. *Vision Research*, *23*, 495–505. doi:10.1016/0042-6989(83)90124-4
- Swift, D. J., & Smith, R. A. (1984). An inherent nonlinearity in near-threshold contrast detection. *Vision Research*, *24*, 977–978. doi:10.1016/0042-6989(84)90073-7
- Watson, A. B. (1986). Temporal sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 6-1–6-43). New York, NY: Wiley.