

Modeling Multisensory Processes in Saccadic Responses: Time-Window-of-Integration Model

Adele Diederich

Hans Colonius

Jacobs University Bremen

Oldenburg University

May 14, 2009

Adele Diederich (*corresponding author*)
School of Humanities and Social Sciences
Jacobs University Bremen
P.O. Box 750 561
D-28725 Bremen
Germany

Phone: +49 421 200 3431

Fax: +49 421 200 3303

email: a.diederich@jacobs-university.de

Hans Colonius
Department of Psychology
Oldenburg University
P.O. Box 2503
D-26111 Oldenburg
Germany

+49 441 798 5158

+49 441 798 5162

hans.colonius@uni-oldenburg.de

1 Summary

Multisensory research within experimental psychology has led to the emergence of a number of lawful relations between response speed and various empirical conditions of the experimental setup (spatiotemporal stimulus configuration, intensity, number of modalities involved, type of instruction, and so forth). This chapter presents a conceptual framework to account for the effects of crossmodal stimulation on response speed. While our framework applies to measures of crossmodal response speed in general, here we focus on modeling *saccadic reaction time* as a measure of reflex-like orientation performance to crossmodal stimuli.

The central postulate is the existence of a critical "time window of integration" (TWIN) controlling the combination of information from different modalities. It is demonstrated that a few basic assumptions about this timing mechanism imply a remarkable number of empirically testable predictions. After introducing a general version of the TWIN model framework, we present various specifications and extensions of the original model that are geared toward more specific experimental paradigms. Our emphasis will be on predictions and empirical testability of these model versions, but for experimental data we refer the reader to the original literature.

2 Multisensory Processes Measured through Response Time

For more than 150 years, response time (RT) has been employed in experimental psychology as a ubiquitous measure to investigate hypotheses about the mental and motor processes involved in simple cognitive tasks (Van Zandt, 2002). Interpreting RT data, in the context of some specific experimental paradigm, is subtle and requires a high level of technical skills. Fortunately, over the years many sophisticated mathematical and statistical methods for response time analysis and corresponding processing models have been developed (Luce, 1986; Schweickert et al., in press). One reason for the sustained popularity of RT as a

measure of mental processes may be the simple fact that these processes always have to unfold over time. A similar rationale, of course, is valid for other methods developed to investigate mental processes, like electrophysiological and related brain-imaging techniques, and it may be one reason why we are currently witnessing some transfer of concepts and techniques from RT analysis into these domains (e.g., Sternberg 2001). Here, we focus on the early, dynamic aspects of simultaneously processing crossmodal stimuli – combinations of vision, audition, and touch – as they are revealed by a quantitative stochastic analysis of response times.

One of the first psychological studies on crossmodal interaction employing RT to measure the effect of combining stimuli from different modalities and of varying their intensities is the classic paper by Todd (1912). A central finding, supported by subsequent research, is that the occurrence of crossmodal effects critically depends on the *temporal* arrangement of the stimulus configuration. For example, the speed-up of response time to a visual stimulus resulting from presenting, say, an accessory auditory stimulus typically becomes most pronounced when the visual stimulus precedes the auditory by an interval that equals the difference in RT between response to the visual alone and the auditory alone (Hershenson, 1962). The rising interest in multisensory research in experimental psychology over the last 20 years has led to the emergence of a number of lawful relations between response speed, on the one hand, and properties of the experimental setting like (i) spatiotemporal stimulus configuration, (ii) stimulus intensity levels, (iii) number of modalities involved, (iv) type of instruction, (v) semantic congruity, on the other. In the following, rather than reviewing the abundance of empirical results, we present a modeling framework within which a number of specific quantitative models have been developed and tested. Although such models can certainly not reflect the full complexity of the underlying multisensory processes, their predictions are sufficiently specific to be rigorously tested through experiments.

For a long time, the ubiquitous mode of assessing response speed has been to measure the time it takes to press a button, or to release it, by moving a finger or foot. With the advance of modern eye movement registration techniques, the measurement of gaze shifts has become an important additional technique to assess multisensory effects. In particular *saccadic* reaction time, i.e., the time from the presentation of a target stimulus to the beginning of the eye

movement, is ideally suited for studying both the temporal and spatial rules of multisensory integration. Although participants can be asked to move their eyes to either visual, auditory, or somatosensory targets, due to the fact that the ocular system is geared to the visual system, the saccadic RT characteristics will be specific to each modality. For example, it is well known that saccades to visual targets have a higher level of accuracy than those to auditory or somatosensory stimuli. Furthermore, as the superior colliculus is an important site of oculomotor control (e.g., Munoz and Wurtz, 1995), measuring saccadic responses is an obvious choice for studying the behavioral consequences of multisensory integration.

3 Time-Window-of-Integration (TWIN) Modeling

We introduce a conceptual framework to account for the effects of crossmodal stimulation as measured by changes in response speed¹. The central postulate is the existence of a critical "time window of integration" (TWIN) controlling the integration of information from different modalities. The starting idea simply is that, say, a visual and an auditory stimulus must not be presented too far away from each other in time for bimodal integration to occur. As we will show, this seemingly innocuous assumption has a number of nontrivial consequences that any multisensory integration model of response speed has to satisfy. Most prominently, it imposes a process consisting of –at least– two serial stages: one early stage, before the outcome of the time window check has occurred, and a later one, where the outcome of the check may affect further processing.

While the TWIN framework applies to measures of crossmodal response speed in general, the focus is on modeling saccadic reaction time. First, a general version of the TWIN model and its predictions, introduced in Colonius & Diederich (2004), will be described. Subsequently, we present various extensions of the original model which are geared toward more specific experimental paradigms. Our emphasis will again be on the predictions and empirical testability of these model versions but, due to space limitations, no experimental data will be presented here.

¹See the Conclusion section for possible extensions to other measures of performance.

3.1 Basic Assumptions

A classic explanation for a speed-up of responses to crossmodal stimuli is that subjects are merely responding to the first stimulus detected. Taking these detection times to be random variables and glossing over some technical details, observed reaction time would then become the minimum of the reaction times to the visual, auditory, or tactile signal leading to a purely statistical facilitation effect (*aka* probability summation) in response speed (Raab, 1962). Over time, numerous studies have shown that this race model is not sufficient to explain the observed speedup in saccadic reaction time (Harrington & Peck, 1998; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Hughes et al., 1998; Corneil & Munoz, 1996; Arndt & Colonius, 2003). Using Millers inequality as a benchmark test (cf. Colonius & Diederich, 2006; Miller, 1982), saccadic responses to bimodal stimuli have been found to be faster than predicted by statistical facilitation, in particular, when the stimuli were spatially aligned. Moreover, in the race model there is no natural explanation for the decrease in facilitation observed with variations in many crossmodal stimulus properties, e.g, increasing spatial disparity between the stimuli.

Nevertheless, the initial anatomic separation of the afferent pathways for different sensory modalities suggests that an early stage of peripheral processing exists, during which no intermodal interaction may occur. For example, a study by Whitchurch and Takahashi (2006) collecting (head) saccadic reaction times in the barn owl lends support to the notion of a race between early visual and auditory processes depending on the relative intensity levels of the stimuli. In particular, their data suggest that the faster modality initiates the saccade, while the slower modality remains available to refine saccade trajectory. Thus, there are good reasons for retaining the construct of an – albeit very peripheral – race mechanism.

Even under invariant experimental conditions, observed responses typically vary from one trial to the next, presumably due to an inherent variability of the underlying neural processes in both ascending and descending pathways. In analogy to the classic race model, this is taken into account in the time-window-of-integration (TWIN) framework by assuming any processing duration to be a random variable. In particular, the peripheral processing times for visual, auditory, and somatosensory stimuli are assumed to be *stochastically independent*

random variables. This leads to the first postulate of the TWIN model:

(B1) **First Stage Assumption:** The first stage consists in a (stochastically independent) race among the peripheral processes in the visual, auditory, and/or somatosensory pathways triggered by a crossmodal stimulus complex.

The existence of a critical "spatiotemporal window" for multisensory integration to occur has been suggested by several authors, based on both neurophysiological and behavioral findings in humans, monkey, and cat (e.g., Bell et al., 2005; Meredith, 2002; Corneil et al., 2002; Meredith et al., 1987; see Navarra et al. 2005 for a recent behavioral study). This integration may manifest itself in the form of an increased firing rate of a multisensory neuron (relative to unimodal stimulation), an acceleration of saccadic reaction time (Frens et al., 1995; Diederich et al., 2003), an effective audiovisual speech integration (Van Wassenhove et al., 2007), or in an improved or degraded judgment of temporal order of bimodal stimulus pairs (cf. Spence and Squire, 2003).

One of the basic tenets of the TWIN framework, however, is the priority of temporal proximity over any other type of proximity: Rather than assuming a joint *spatiotemporal* window of integration permitting interaction to occur only for both spatially and temporally neighboring stimuli, the TWIN model allows for crossmodal interaction to occur, for example, even for spatially rather distant stimuli of different modalities as long as they fall within the time window.

(B2) **Time-Window-of-Integration Assumption:** Multisensory integration occurs *only* if the peripheral processes of the first stage all terminate within a given temporal interval, the "time window of integration".

In other words, a visual and an auditory stimulus may occur at the same spatial location, or the lip movements of a speaker may be perfectly consistent with the utterance, no intersensory interaction effect will be possible if the data from the two sensory channels are registered too distant from each other in time. Thus, the window acts like a *filter* determining whether afferent information delivered from different sensory organs is registered close

enough in time to allow for multisensory integration. Note that passing the filter is a necessary, but not sufficient, condition for multisensory integration to occur. The reason is that the amount of multisensory integration also depends on other aspects of the stimulus set like the spatial configuration of the stimuli. For example, response depression may occur with nearly simultaneous but distant stimuli, making it easier for the organism to focus attention on the more important event. In other cases, multisensory integration may fail to occur—despite near-simultaneity of the unisensory events—because the a-priori probability for a crossmodal event is very small (e.g., Körding et al. 2007).

While the priority of temporal proximity seems to afford more flexibility for an organism in a complex environment, the next assumption delimits the role of temporal proximity to the first processing stage:

(B3) Assumption of Temporal Separability: The amount of interaction manifesting itself in an increase or decrease of second stage processing time, is a function of crossmodal stimulus features, but it does not depend on the presentation asynchrony (SOA) of the stimuli.

This assumption is based on a distinction between intra- and crossmodal stimulus properties, where the properties may refer to both subjective and physical properties. *Crossmodal* properties are defined when stimuli of more than one modality are present, like spatial distance of target to nontarget, or subjective similarity between stimuli of different modalities. *Intramodal* properties, on the other hand, refer to properties definable for a single stimulus, no matter whether this property is definable in all modalities (like intensity) or in only one modality (like wavelength for color or frequency for pitch). Intramodal properties can affect the outcome of the race in the first stage and, thereby, the probability of an interaction. Crossmodal properties may affect the amount of crossmodal interaction occurring in the second stage. Note that crossmodal features cannot influence first stage processing time, since the stimuli are still being processed in separate pathways.

(B4) Second Stage Assumption: The second stage comprises all processes following the first stage including preparation and execution of a response.

The assumption of only two stages is certainly an oversimplification. Note, however, that the second stage is defined here by default: it includes all subsequent, possibly overlapping, processes that are not part of the peripheral processes in the first stage (for a similar approach, see Van Opstal and Munoz 2004). Thus, the TWIN model retains the classic notion of a race mechanism as an explanation for crossmodal interaction but restricts it to the very first stage of stimulus processing.

3.2 Quantifying multisensory integration in the TWIN Model

In order to derive empirically testable predictions from the TWIN framework, its assumptions must be put into a more precise form. According to the two-stage assumption, total saccadic reaction time in the crossmodal condition can be written as a sum of two nonnegative random variables defined on a common probability space:

$$RT_{crossmodal} = S_1 + S_2, \quad (1)$$

where S_1 and S_2 refer to first and second stage processing time, respectively (a base time would also be subsumed under S_2). Let I denote the event that multisensory integration occurs, having probability $P(I)$. For the expected reaction time in the crossmodal condition then follows:

$$\begin{aligned} \mathbb{E}[RT_{crossmodal}] &= \mathbb{E}[S_1] + \mathbb{E}[S_2] \\ &= \mathbb{E}[S_1] + P(I) \cdot \mathbb{E}[S_2|I] + (1 - P(I)) \cdot \mathbb{E}[S_2|I^c] \\ &= \mathbb{E}[S_1] + \mathbb{E}[S_2|I^c] - P(I) \cdot (\mathbb{E}[S_2|I^c] - \mathbb{E}[S_2|I]), \end{aligned}$$

where $\mathbb{E}[S_2|I]$ and $\mathbb{E}[S_2|I^c]$ denote the expected second stage processing time conditioned on interaction occurring (I) or not occurring (I^c), respectively. Putting $\Delta \equiv \mathbb{E}[S_2|I^c] - \mathbb{E}[S_2|I]$, this becomes

$$\mathbb{E}[RT_{crossmodal}] = \mathbb{E}[S_1] + \mathbb{E}[S_2|I^c] - P(I) \cdot \Delta. \quad (2)$$

That is, mean RT to crossmodal stimuli is the sum of mean RT of the first stage processing time, mean RT of the second stage processing when no interaction occurs, and the term $P(I) \cdot \Delta$ which is a measure of the expected amount of intersensory interaction in the second stage with positive Δ values corresponding to facilitation, negative ones to inhibition.

This factorization of expected intersensory interaction into the probability of interaction $P(I)$ and the amount and sign of interaction (Δ) is an important feature of the TWIN model. According to assumptions (B1) to (B4), the first factor, $P(I)$, depends on the temporal configuration of the stimuli (stimulus onset asynchrony, SOA), whereas the second factor, Δ , depends on non-temporal aspects, in particular their spatial configuration. Note that this *separation of temporal and non-temporal factors* is in accordance with the definition of the window of integration: the incidence of multisensory integration hinges upon the stimuli to occur in temporal proximity, whereas the amount and sign of interaction (Δ) is modulated by non-temporal aspects like semantic congruity or spatial proximity reaching, in the latter case, from enhancement for neighboring stimuli to possible inhibition for distant stimuli (cf. Diederich and Colonius 2007b).

3.3 Some general predictions of TWIN

In the section following this one, more specific assumptions on first stage processing time S_1 and probability of interaction $P(I)$ will be introduced in order to derive detailed quantitative predictions for specific experimental crossmodal paradigms. Nonetheless, even at the general level of the framework introduced so far, a number of qualitative empirical predictions of TWIN are possible.

SOA effects. The amount of crossmodal interaction should depend on the stimulus onset asynchrony (SOA) between the stimuli, since the probability of integration $P(I)$ changes with SOA. Let us assume that two stimuli from different modalities differ considerably in their peripheral processing times. If the faster stimulus is delayed (in terms of SOA) so that the arrival times of both stimuli have a high probability of falling into the window of integration, then the amount of crossmodal interaction should be largest for that value of SOA (see, e.g.,

Frens et al., 1995; Colonius & Arndt, 2001).

Intensity effects. Stimuli of high intensity have relatively fast peripheral processing times. Therefore, for example, if a stimulus from one modality has a high intensity compared to a stimulus from the other modality, the chance that both peripheral processes terminate within the time window will be small, assuming simultaneous stimulus presentations. The resulting low value of $P(I)$ is in line with the empirical observation that a very strong signal will effectively rule out any further reduction of saccadic RT by adding a stimulus from another modality (e.g., Corneil, Wanrooij, Munoz, & Van Opstal, 2002).

Crossmodal effects. The amount of multisensory integration (Δ) and its sign (facilitation or inhibition) occurring in the second stage depend on crossmodal features of the stimulus set, for example spatial disparity and laterality (laterality here refers to whether or not all stimuli appear in the same hemisphere). Crossmodal features cannot have an influence on first stage processing time since the modalities are yet being processed in separate pathways. Conversely, parameter Δ not depending on SOA cannot change its sign as a function of SOA and, therefore, the model cannot simultaneously predict facilitation to occur for some SOA values and inhibition for others. Some empirical evidence against this prediction has been observed (Diederich and Colonius, 2008a).

In the classic race model, the addition of a stimulus from a modality not yet present will increase (or, at least, not decrease) the amount of response facilitation. This follows from the fact that –even without assuming stochastic independence– the probability of the fastest of several processes terminating processing before time t will increase with the number of "racers" (e.g., Colonius and Vorberg 1994) . In the case of TWIN, both facilitation and inhibition are possible under certain conditions as follows:

Number of modalities effect. The addition of a stimulus from a modality not yet present will increase (or, at least, not decrease) the expected amount of interaction if the added stimulus is not "too fast" and the time window is not "too small". The latter restrictions are meant to guarantee that the added stimulus will fall into the time window, thereby increasing the probability of interaction to occur.

4 TWIN Models for Specific Paradigms: Assumptions and Predictions

In a crossmodal experimental paradigm, the individual modalities may either be treated as being on an equal footing, or one modality may be singled out as target modality, while stimuli from the remaining modalities may be ignored by the participant as nontargets. Crossmodal effects are assessed in different ways, depending on task instruction. As shown below, the TWIN model can take these different paradigms into account simply by modifying the conditions that lead to an opening of the time window.

4.1 Measuring Crossmodal Effects in Focused Attention and Redundant Target Paradigms

In the *redundant target* (aka *divided-attention*) paradigm (RTP), stimuli from different modalities are presented simultaneously or with certain stimulus onset asynchrony (SOA), and the participant is instructed to respond to the stimulus detected first. Typically, the time to respond in the crossmodal condition is faster than in either of the unimodal conditions. In the *focused attention* paradigm (FAP), crossmodal stimulus sets are presented in the same manner but now participants are instructed to respond only to the onset of a stimulus from a specifically defined target modality, such as the visual, and to ignore the remaining nontarget stimulus, the tactile or the auditory, say. In the latter setting, when a stimulus of a nontarget modality, a tone, say, appears before the visual target at some spatial disparity, there is no *overt* response to the tone if the participant is following the task instructions. Nevertheless, the nontarget stimulus has been shown to modulate the saccadic response to the target: Depending on the exact spatiotemporal configuration of target and nontarget, the effect can be a speed-up or an inhibition of saccadic RT (see, e.g., Amlôt et al. 2003, Diederich and Colonius 2007b), and the saccadic trajectory can be affected as well (Doyle and Walker 2002).

Some striking similarities to human data have been found in a detection task utilizing both paradigms. Stein, Huneycutt, and Meredith (1988) trained cats to orient to visual or auditory stimuli, or both. In one paradigm, the target was a visual stimulus (a dimly illuminating LED) and the animal learned that although an auditory stimulus (a brief, low-intensity broad-band noise) would be presented periodically, responses to it would never be rewarded, and the cats learned to "ignore" it (focused attention paradigm). Visual-auditory stimuli were always presented spatially coincident, but their location varied from trial to trial. The weak visual stimulus was difficult to detect and the cats' performance was below 50 % correct detection. However, combining the visual stimulus with the neutral auditory stimulus markedly enhanced performance, regardless of their position. A similar result was obtained when animals learned that both stimuli were potential targets (redundant target paradigm). In a separate experiment in which the visual and the (neutral) auditory stimuli were spatially disparate, however, performance was significantly worse than when the visual stimulus was presented alone (cf. Stein et al. 2004).

A common method to assess the amount of crossmodal interaction is to use a measure that relates mean RT in crossmodal conditions to that in the unimodal condition. The following definitions quantify the percent RT enhancement in analogy to a measure proposed for measuring multisensory enhancement in neural responses (cf. Meredith and Stein, 1986a; Anastasio et al. 2000; Colonius and Diederich, 2002; Diederich and Colonius 2004a,b). For visual, auditory, and visual-auditory stimuli with observed mean (saccadic or manual) reaction time RT_V , RT_A and RT_{VA} , respectively, and $SOA = \tau$, the multisensory response enhancement (MRE) for the redundant target task is defined as

$$\text{MRE}_{RTP} = \frac{\min(RT_V, RT_A + \tau) - RT_{VA, \tau}}{\min(RT_V, RT_A + \tau)} \cdot 100, \quad (3)$$

where $RT_{AV, \tau}$ refers to observed mean RT to the bimodal stimulus with $SOA = \tau$. For the focused attention task, MRE is defined as

$$\text{MRE}_{FAP} = \frac{RT_V - RT_{VA, \tau}}{RT_V} \cdot 100, \quad (4)$$

assuming vision as target modality.

4.2 TWIN Model for Focused Attention Paradigm

TWIN is adapted to the focused attention task through replacing the original time-window-of-integration assumption (B2) by

(B2-FAP) Time-Window-of-Integration Assumption: In the focused attention paradigm, crossmodal interaction occurs only if (i) a *nontarget* stimulus wins the race in the first stage, opening the time window of integration such that (ii) the termination of the target peripheral process falls in the window. The duration of the time window is a constant.

The idea here is that the winning nontarget will keep the saccadic system in a state of heightened reactivity such that the upcoming target stimulus, if it falls into the time window, will trigger crossmodal interaction. At the neural level, this may correspond to a gradual inhibition of fixation neurons (in superior colliculus) and/or omnipause neurons (in midline pontine brain stem). In the case of the target being the winner, no discernible effect on saccadic RT is predicted, like in the unimodal situation.

The race in the first stage of the model is made explicit by assigning statistically independent, nonnegative random variables V and A to the peripheral processing times for a visual target and an auditory nontarget stimulus, say, respectively. With τ as SOA value and ω as integration window width parameter, assumption (B2-FAP) amounts to the event that multisensory integration occurs, I_{FAP} , being

$$I_{FAP} = \{A + \tau < V < A + \tau + \omega\}.$$

Thus, the probability of integration to occur, $P(I_{FAP})$, is a function of both τ and ω , and it can be determined numerically once the distribution functions of A and V have been specified.

Expected reaction time in the bimodal condition then is (cf. Eq. 2)

$$E[RT_{VA,\tau}] = E[V] + E[S_2|I_{FAP}^c] - P(I_{FAP}) \cdot \Delta. \quad (5)$$

No interaction is possible in the unimodal condition. Thus, expected reaction time for the visual (target) stimulus condition is

$$E[RT_V] = E[V] + E[S_2|I_{FAP}^c]. \quad (6)$$

Note that in the focused attention task, the first stage duration is defined as the time it takes to process the (visual) target stimulus, $E[V]$. Crossmodal interaction (CI) is defined as difference between mean RT to the unimodal and crossmodal stimuli, i.e.,

$$CI \equiv E[RT_V] - E[RT_{VA,\tau}] = P(I_{FAP}) \cdot \Delta. \quad (7)$$

Thus, the separation of temporal and non-temporal factors expressed in the above equation for the observable CI is directly inherited from assumptions (B4) and (B2-FAP).

4.2.1 TWIN Predictions for Focused Attention Paradigm

The integration assumption (B2-FAP) permits further specification of TWIN's general predictions of Section 3.3. From a model testing point of view, it is a clear strength of the TWIN framework that it allows for numerous qualitative predictions without having to specify the probability distributions for the random processing times. Thus, a violation of any one of these predictions cannot be attributed to an inappropriate choice of the distributions but may point to a more fundamental inadequacy of one or, possibly, several model assumptions. For a quantitative fit to an observed set of data, however, some distributional assumptions are required. In the parametric version of TWIN, all peripheral processing times are assumed to be exponentially distributed (cf. Colonius and Diederich 2004). This choice is made mainly for computational simplicity: calculating the probability of integration, $P(I_{FAP})$, is straightforward, and the exponential distribution is characterized by a single quantity, the *intensity parameter* λ (see appendix). As long as predictions are limited to the level of means, no specific assumptions about the distribution of processing times in the second stage are necessary (but see Section 6).

Next, we demonstrate how the focused attention context leads to more specific empiri-

cally testable predictions of TWIN. Predictions relying on the parametric TWIN version are postponed to the final part of this section. If not specifically mentioned otherwise, we always assume nonnegative Δ values in the following elaborations.

SOA effects. When the nontarget is presented very late relative to the target (large positive SOA), its chance of winning the race against the target and thus opening the window of integration become very small. When it is presented rather early (large negative SOA), it is likely to win the race and to open the window, but the window may be closed by the time the target arrives. Again, the probability of integration, $P(I_{FAP})$, is small. Therefore, the largest probability of integration is expected for some mid-range SOA values. Although $P(I_{FAP})$ is unobservable, it should leave its mark on a wellknown observable measure, i.e., multisensory response enhancement (MRE). In fact, MRE_{FAP} , defined in (Eq. 4), as a function of SOA, should have the same form as $P(I_{FAP})$, scaled only by some constant:

$$\begin{aligned} MRE_{FAP} &= \frac{(RT_V - RT_{VA, \tau})}{RT_V} \cdot 100 \\ &= \frac{P(I_{FAP}) \cdot \Delta}{RT_V} \cdot 100 \\ &= P(I_{FAP}) \cdot \Delta \cdot \text{const.} \end{aligned} \tag{8}$$

Intensity effects. Increasing the intensity of the visual stimulus will speed up visual peripheral processing (up to some minimum level) thereby increasing the chance for the visual target to win the race. Thus, the probability that the window of integration opens decreases, predicting less multisensory integration. Increasing the intensity of the nontarget auditory stimulus, on the other hand, leads to the opposite prediction: the auditory stimulus will have a better chance to win the race and to open the window of integration, hence predicting more multisensory integration to occur on average. Two further distinctions can be made. For large negative SOA, i.e., when the auditory nontarget arrives very early, further increasing the auditory intensity makes it more likely for the time window of integration to close before the target arrives and therefore results in a lower $P(I_{FAP})$ value. For smaller negative SOA, however, i.e., when the nontarget is presented shortly before the target, increasing the auditory intensity improves its chances to win against the target and to open the window. Given the complexity of these intensity effects, however, more specific quantitative predic-

tions will require some distributional assumptions for the first stage processing times (see below). Alternatively, it may be feasible to adapt the "double factorial paradigm" developed in Townsend and Nozawa (1995) to analyze predictions when the effects of both targets and nontargets presented at two different intensities levels are observed.

Crossmodal effects. If target and nontarget are presented in two distinct crossmodal conditions, one would expect parameter Δ to take on two different values. For example, for two spatial conditions, ipsilateral and contralateral, the values could be Δ_i and Δ_c say. Subtracting the corresponding crossmodal interaction terms then gives (cf. Eq. 7)

$$CI_i - CI_c = P(I_{FAP}) \cdot (\Delta_i - \Delta_c), \quad (9)$$

an expression that should again yield the same qualitative behavior, as a function of SOA, as $P(I_{FAP})$. In a similar vein, one can capitalize on the factorization of expected crossmodal interaction if some additional experimental factor affecting Δ , but not $P(I_{FAP})$, is available. In Colonius et al. (2009), an auditory background masker stimulus, presented at increasing intensity levels, was hypothesized to simultaneously increase Δ_c and decrease Δ_i . The ratio of CIs in both configurations,

$$\frac{CI_i}{CI_c} = \frac{P(I_{FAP}) \cdot \Delta_i}{P(I_{FAP}) \cdot \Delta_c} = \frac{\Delta_i}{\Delta_c}, \quad (10)$$

should then remain invariant across SOA values, with a separate value for each level of the masker.

Number of nontargets effects. For crossmodal interaction to occur in the focused attention task, it is necessary that the nontarget process wins the race in the first stage. With two or more nontargets entering the race, the probability of one of them winning against the target process increases and, therefore, the probability of opening the window of integration increases with the number of nontargets present. In this case, there are even two different ways of utilizing the factorization of CI, both requiring the existence of two crossmodal conditions with two different Δ parameters (spatial or other). The first test is analogous to the previous one. Since the number of nontargets affects $P(I_{FAP})$ only, the ratio in Eq. 10 should be the same whether it is computed from conditions with one or two

nontargets. The second test results from taking the ratio of CI based on one nontarget, C_1 , over CI based on two nontargets, C_2 . Since Δ should not be affected by the number of nontargets, the ratio

$$\frac{CI_1}{CI_2} = \frac{P_1(I_{FAP}) \cdot \Delta}{P_2(I_{FAP}) \cdot \Delta} = \frac{P_1(I_{FAP})}{P_2(I_{FAP})}, \quad (11)$$

where P_1 and P_2 refer to the probability of opening the window under one or two nontargets, respectively, should be the same, no matter from which one of the two crossmodal conditions it was computed. In Diederich and Colonius (2007a), neither of these tests revealed evidence against these TWIN predictions.

SOA and intensity effects predicted by a parametric TWIN version. Assuming exponential distributions for the peripheral processing times, the intensity parameter for the visual modality is set to $1/\lambda_V = 50$ [ms] and to $1/\lambda_A = 10, 30, 70,$ or 90 [ms] for the auditory nontarget. Quantitative predictions of TWIN for focused attention are shown in the left of Figure 1. Panels 1 and 2 show mean RT and $P(I_{FAP})$ as a function of SOA for the various intensities of the auditory nontarget. Note that two intensities result in faster and two in lower mean RT compared to mean unimodal RT to the visual target. Here, the parameter for second stage processing time when no integration occurs, μ , was set to 100 ms. The time window of integration was set to 200 ms. The parameter for multisensory integration was set to $\Delta_i = 20$ ms for bimodal stimuli presented ipsilaterally, implying a facilitation effect. Note that neither λ_V nor μ are directly observable, but the sum of the peripheral and central processing time for the visual target stimulus constitutes a prediction for unimodal mean saccadic RT:

$$E[RT_V] = \frac{1}{\lambda_V} + \mu,$$

which, for the present example, is $50 \text{ ms} + 100 \text{ ms} = 150 \text{ ms}$. The dashed (blue) line and the dotted (black) line show the bimodal RT predictions for the auditory nontargets with the highest and lowest intensity, respectively.

No fits to empirical data sets are presented here, but good support of TWIN has been found so far (see, e.g., Diederich and Colonius, 2007a,b, 2008a,b). Close correspondence between data and model prediction, however, is not the only aspect to consider. Importantly, the pattern of parameter values estimated for a given experimental setting should suggest

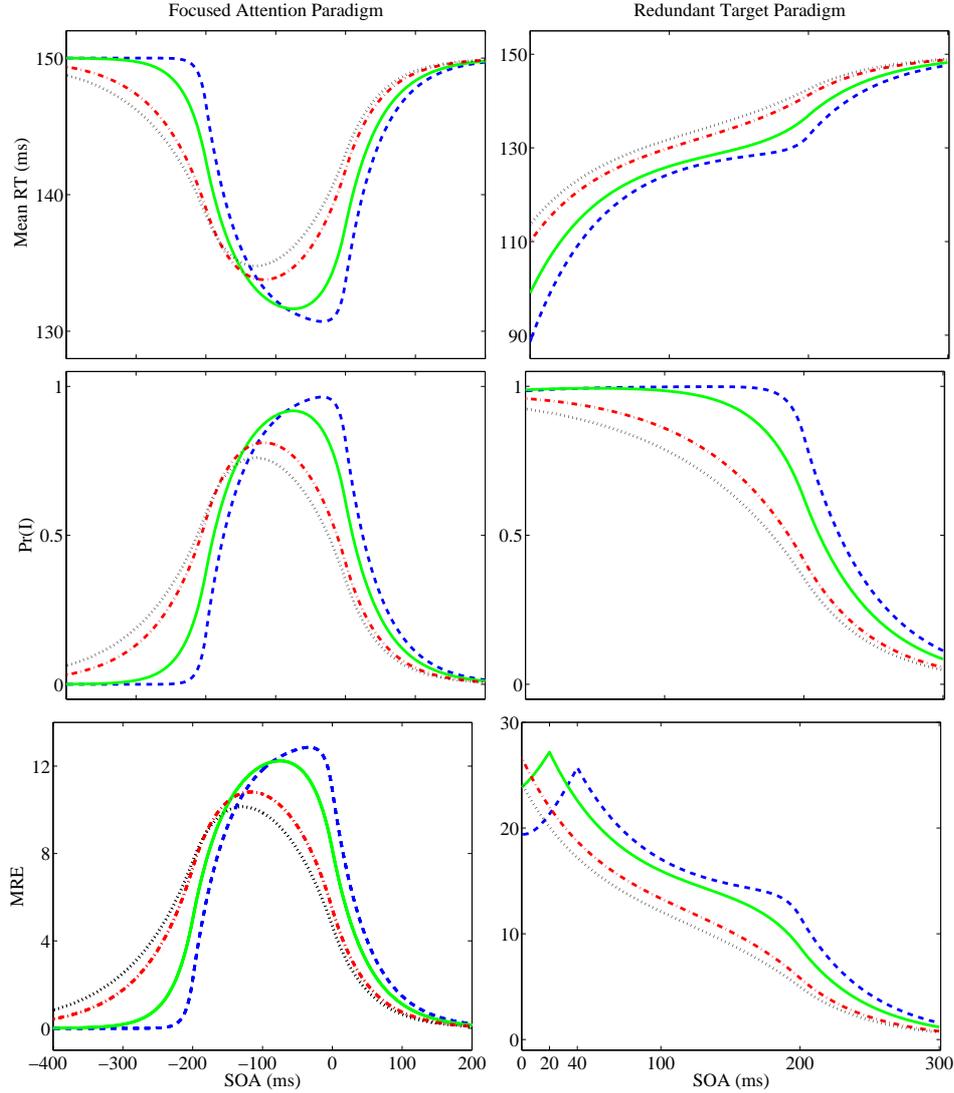


Figure 1: TWIN predictions for focused attention paradigm (FAP) (left panels) and redundant target paradigm (RTP) (right panels). The parameters in both paradigms are chosen to be identical. Mean RT for the visual stimulus is 150 ms ($1/\lambda_V = 50, \mu = 100$). The peripheral processing times for the auditory stimuli are $1/\lambda_A = 10$ [ms] (blue dashed line); $1/\lambda_A = 30$ [ms] (green solid); $1/\lambda_A = 70$ [ms] (red dash-dotted); and $1/\lambda_A = 90$ [ms] (black dotted). Interaction parameter is $\Delta = 20$ [ms].

a meaningful interpretation. For example, increasing stimulus intensity are reflected in a decrease of the corresponding λ parameters, assuming higher intensities to lead to faster peripheral processing times (at least, within certain limits). Or, in the study with an auditory background masker (Colonius et al., 2009), the crossmodal interaction parameter (Δ) was a decreasing or increasing function of masker level for the contralateral or ipsilateral condition,

respectively, as predicted.

4.3 TWIN Model for Redundant Target Paradigm

TWIN is adapted to the redundant target task through replacing the original time-window-of-integration assumption (B2) by

(B2-RTP) Time-Window-of-Integration Assumption: In the redundant target paradigm, (i) the window of integration is opened by whichever stimulus wins the race in the first stage, and (ii) crossmodal interaction occurs if the termination of the peripheral process of a stimulus of another modality falls within the window. The duration of the time window is a constant.

Obviously, if stimuli from more than two modalities are presented, the question of a possible additional effect on crossmodal interaction arises. There is both behavioral and neurophysiological evidence for trimodal interaction (e.g., Diederich and Colonius, 2004b; Stein and Meredith 1993), but data from saccadic eye movement recordings seem not yet to be conclusive enough to justify further elaboration of assumption B2-RTP.

To compute the probability of interaction in the redundant target paradigm, $P(I_{RTP})$, we assume that a visual and an auditory stimulus are presented with an SOA equal to τ . Then, either the visual stimulus wins, $V < A + \tau$, or the auditory stimulus wins, $A + \tau < V$; so, in either case, $\min(V, A + \tau) < \max(V, A + \tau)$ and, by assumption B2-RTP,

$$I_{RTP} = \{\max(V, A + \tau) < \min(V, A + \tau) + \omega\}.$$

Thus, the probability of integration to occur is a function of both τ and ω , as before. Expected reaction time in the crossmodal condition is computed as (see Eq. 2):

$$E[RT_{VA,\tau}] = E[\min(V, A + \tau)] + E[S_2|I_{RTP}^c] - P(I_{RTP}) \cdot \Delta. \quad (12)$$

In the redundant target paradigm, first stage duration is determined by the termination time of the winner. This is an important difference to the focused attention situation where first stage duration is defined by the time it takes to process the (visual) target stimulus. Even for a zero probability of interaction, expected reaction time in the bimodal condition is smaller than, or equal to, either of the unimodal stimulus conditions. These are

$$E[RT_V] = E[V] + E[S_2|I_{RTP}^c] \quad (13)$$

and

$$E[RT_A] = E[A] + E[S_2|I_{RTP}^c], \quad (14)$$

due to the fact that in the redundant target version of TWIN, the race in first stage produces a statistical facilitation effect equivalent to the one in the classic race model. Thus, a possible crossmodal enhancement observed in a redundant target task may be due to multisensory integration or statistical facilitation, or both. Moreover, a possible crossmodal inhibition effect may be weakened by simultaneous presence of statistical facilitation in the first stage. Predictions for the redundant target case are less straightforward than for focused attention because the factorization of crossmodal interaction (CI) in the latter is no longer valid. Nevertheless, some general predictions can be made assuming, as before, a multisensory facilitation effect, i.e., $\Delta > 0$.

4.3.1 TWIN Predictions for Redundant Target Paradigm

In this paradigm, both stimuli are on an equal footing and, therefore, negative SOA values need not be introduced. Each SOA value now indicates the time between the stimulus presented first and the one presented second, regardless of modality.

SOA effects. The probability of crossmodal interaction decreases with increasing SOA: The later the second stimulus is presented, the less likely it is to win the race and to open the window of integration; alternatively, if the window has already been opened by the first stimulus, the less likely it is to fall into that window with increasing SOA. For large enough SOA values, mean saccadic RT in the crossmodal condition approaches the mean for the

stimulus presented first .

To fix ideas, we now assume, without loss of generality, that a visual stimulus of constant intensity is presented first and that an auditory stimulus is presented second, or simultaneous with the visual, and at different intensities. Predictions then depend on the *relative* intensity difference between both stimuli. Note that the unimodal means constitute upper bounds for bimodal mean RT.

Intensity effects. For a visual stimulus presented first, increasing the intensity of the auditory stimulus (presented second) increases the amount of facilitation.

SOA and intensity effects predicted by a parametric TWIN version. Figure 1 (right panels) shows the quantitative predictions of TWIN for SOA and intensity variations under exponential distributions for the peripheral processing times. Parameters are the same as for the FAP predictions (left panels). Panels 1 and 2 show mean RT and $P(I)$ as a function of SOA for various intensity levels (λ parameters) of the auditory stimulus. Both panels exhibit the predicted monotonicity in SOA and intensity. The third panel, depicting multisensory response enhancement (MRE), reveals some non-monotonic behavior in both SOA and intensity.

Without going into numerical details, this non-monotonicity of MRE can be seen to be due to a subtle interaction between two mechanism both being involved in the generation of MRE: (i) statistical facilitation occurring in the first stage and (ii) opening of the time window. The former is maximal if presentation of the stimulus processed faster is delayed by an SOA equal to the difference in mean RT in the unimodal stimulus conditions, that is when peripheral processing times are in *physiological synchrony*; for example, if mean RT to an auditory stimulus is 110 ms and mean RT to a visual stimulus is 150 ms, the maximal amount of statistical facilitation is expected when the auditory stimulus is presented $150\text{ ms} - 110\text{ ms} = 40\text{ ms}$ after the visual stimulus. The SOA value being "optimal" for statistical facilitation, however, need not be the one producing the highest probability of opening the time window which was shown to be decreasing with SOA. Moreover, the non-monotonicity in intensity becomes plausible if one realizes that variation in intensity results in a change of mean processing time analogous to an SOA effect: for example, lowering

auditory stimulus intensity has an effect on statistical facilitation and the probability of opening the time window that is comparable to increasing SOA.

4.4 Focused attention vs. redundant target paradigm

Top-down vs. bottom-up. The distinction between redundant target and focused attention paradigm is not only an interesting experimental variation as such but it may also provide an important theoretical aspect. In fact, since in both paradigms physically identical stimuli can be presented under the same spatiotemporal configuration, any differences observed in the corresponding reaction times would have to be due to the instructions being different, thereby pointing to a possible separation of top-down from bottom-up processes in the underlying multisensory integration mechanism.

Probability of integration. Moreover, comparing both paradigms yields some additional insight into the mechanics of TWIN. Note that under equivalent stimulus conditions, $I_{FAP} \subset I_{RTP}$; this relation follows from the observation that

$$I_{FAP} = I_{RTP} \cap \{A + \tau \text{ is the winner of the race}\}.$$

It means that any realization of the peripheral processing times that leads to an opening of the time window under the focused attention instruction also leads to the same event under the redundant target instruction. Thus, the probability of integration under redundant target instructions can not be smaller than that under focused attention instruction: $P(I_{FAP}) \leq P(I_{RTP})$, given identical stimulus conditions (see also Fig. 1).

Inverse effectiveness. It is instructive to consider the effect of varying stimulus intensity in both paradigms when both stimuli are presented simultaneously (SOA= 0) and at intensity levels producing the same mean peripheral speed, i.e., with the same intensity parameters, $\lambda_V = \lambda_A$. Assuming exponential distributions, Figure 2 depicts probability of integration (upper panels) and multisensory response enhancement (MRE) (lower panels) as a function of time window width (ω) for both paradigms and with each curve presenting a specific intensity level. Probability of integration increases monotonically from zero (for $\omega = 0$)

towards 0.5 for the focused attention, and towards 1 for the redundant target paradigm. For the former, probability of integration cannot surpass 0.5 because, for any given window width, the target process has the same chance of winning as the nontarget process under the given λ parameters. For both paradigms, $P(I)$, as a function of ω , is ordered with respect to intensity level: it increases monotonically with the mean processing time of both stimuli² (upper panels of Figure 2). The same ordering is found for MRE in the focused attention paradigm; somewhat surprisingly, however, the ordering is reversed for MRE in the redundant target paradigm: increasing intensity implies less enhancement, i.e., it exhibits the "inverse effectiveness" property often reported in empirical studies (Stein and Meredith 1993; Rowland and Stein 2008). Similar to the above discussion of intensity effects for RTP, this is due to an interaction generated by increasing intensity: it weakens statistical facilitation in first stage processing but simultaneously increases the probability of integration.

5 TWIN model for Focused Attention: Including a Warning Mechanism

Although estimates for the time window of integration vary somewhat across subjects and task specifics, a 200 ms width showed up in several studies (e.g., Eimer 2001; Sinclair and Hammond 2009). In a focused attention task, when the nontarget occurs at an early point in time (i.e., 200 ms or more before the target), a substantial decrease of RT compared to the unimodal condition has been observed in Diederich and Colonius (2007a). This decrease, however, no longer depended on whether target and nontarget appeared at ipsi- or contralateral positions, thus supporting the hypothesis that the nontarget plays the role of a *spatially unspecific* alerting cue, or warning signal, for the upcoming target whenever the SOA is large enough.

The hypothesis of increased crossmodal processing triggered by an alerting cue had already been advanced in Nickerson (1973) who called it "preparation enhancement". In the eye movement literature, the effects of a warning signal have been studied primarily in the

²This is due to a property of the exponential distribution: mean and standard deviation are identical.

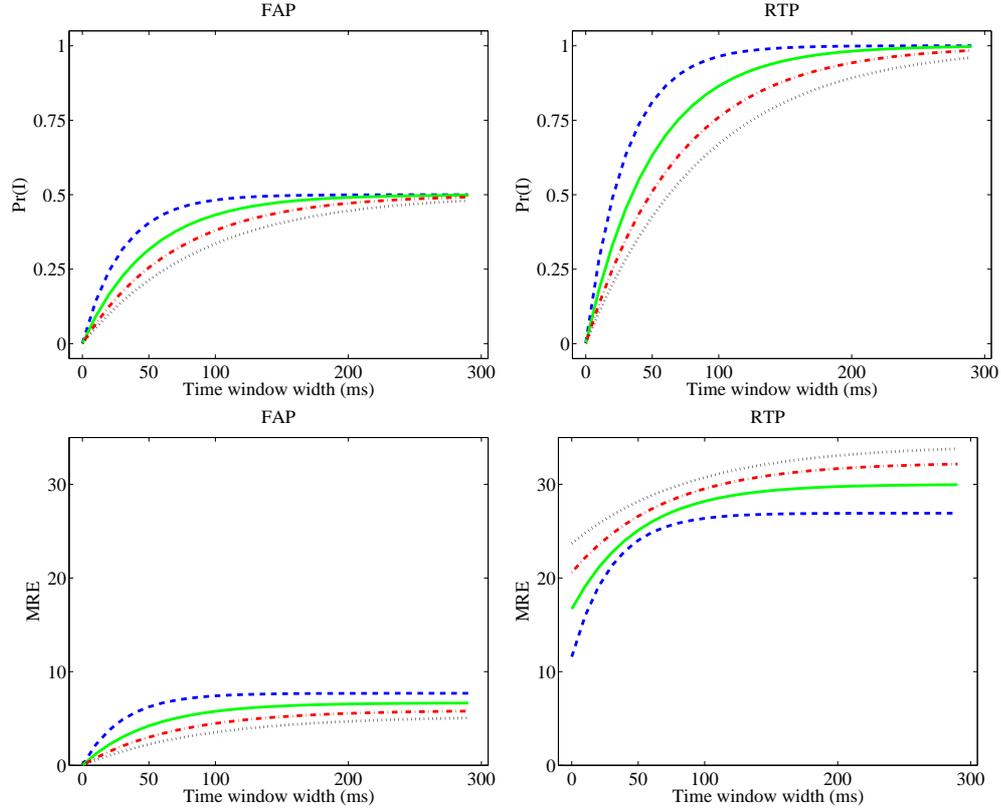


Figure 2: TWIN predictions for focused attention paradigm (FAP) (left panels) and redundant target paradigm (RTP) (right panels) as a function of time window width (ω) at SOA = 0. Upper panels depict probability of integration $P(I)$, the lower panels multisensory response enhancement (MRE). Each curve corresponds to a specific intensity parameter of the stimuli. The peripheral processing times for the auditory and visual stimuli are $1/\lambda_A = 1/\lambda_V$ equal to 30 (blue dashed line); 50 (green solid); 70 (red dash-dotted); and 90 (black dotted). Mean second stage processing time is $\mu = 100$. Interaction parameter is $\Delta = 20$. [all values in ms]

context of explaining the "gap effect", i.e., the latency to initiate a saccade to an eccentric target is reduced by extinguishing the fixation stimulus about 200 ms prior to target onset (Reuter-Lorenz et al. 1991; Klein and Kingston 1993). An early study on the effect of auditory or visual warning signals on saccade latency, but without considering multisensory integration effects, was conducted by Ross and Ross (1981).

Here, the dual role of the nontarget – inducing multisensory integration that is governed by the above mentioned spatiotemporal rules, on the one hand, and acting as a spatially unspecific crossmodal warning cue, on the other – will be taken into account by an extension

of TWIN that yields an estimate of the relative contribution of either mechanism for any specific SOA value.

(W) Assumption on warning mechanism: If the nontarget wins the processing race in the first stage by a margin wide enough for the time window of integration to be closed again before arrival of the target, then subsequent processing will be facilitated or inhibited ("warning effect") without dependence on the spatial configuration of the stimuli³.

The time margin by which the nontarget may win against the target will be called *headstart* denoted as γ . The assumption stipulates that the headstart is at least as large as the width of the time window for a warning effect to occur. That is, the warning mechanism of the nontarget is triggered whenever the nontarget wins the race by a headstart $\gamma \geq \omega \geq 0$. Taking, for concreteness, the auditory as nontarget modality, occurrence of a warning effect corresponds to the event:

$$W = \{A + \tau + \gamma < V\}.$$

The probability of warning to occur, $P(W)$, is a function of both τ and γ . Since $\gamma \geq \omega \geq 0$ this precludes simultaneous occurrence of both warning and multisensory interaction within one and the same trial and, therefore, $P(I \cap W) = 0$ (since no confusion can arise, we write I for I_{FAP} throughout this section). The actual value of the headstart criterion is a parameter to be estimated in fitting the model under Assumption (W).

The expected saccadic reaction time in the crossmodal condition in the TWIN model with warning assumption can then be shown to be:

³In Diederich and Colonius 2008a, an alternative version of this assumptions was considered as well: (Version B) *If the nontarget wins the processing race in the first stage by a wide enough margin, then subsequent processing will in part be facilitated or inhibited without dependence on the spatial configuration of the stimuli.* This version is less restrictive: All that is needed for the nontarget to act as a warning signal is a "large enough" headstart against the target in the race and $P(I \cap W)$ can be larger than 0. Assuming that the effects on RT of the two events I and W , integration and warning, combine additively, it can then be shown that the crossmodal interaction prediction of this model version is captured by the same equation as under the original version, i.e. Eq. 17 below. The only difference is in the order restriction for the parameters, $\gamma \geq \omega$. Up to now, no empirical evidence was found in favor of one of the two versions over the other.

$$\begin{aligned}
E[RT_{crossmodal}] &= E[S_1] + E[S_2] \\
&= E[S_1] + E[S_2|I^c \cap W^c] - P(I) \cdot \{E[S_2|I^c \cap W^c] - E[S_2|I]\} \\
&\quad - P(W) \cdot \{E[S_2|I^c \cap W^c] - E[S_2|W]\},
\end{aligned}$$

where $E[S_2|I]$, $E[S_2|W]$, and $E[S_2|I^c \cap W^c]$ denote the expected second stage processing time conditioned on interaction occurring (I), warning occurring (W), or neither of them occurring ($I^c \cap W^c$), respectively (I^c, W^c stand for the complement of events I, W). Setting

$$\begin{aligned}
\Delta &\equiv E[S_2|I^c \cap W^c] - E[S_2|I] \\
\kappa &\equiv E[S_2|I^c \cap W^c] - E[S_2|W]
\end{aligned}$$

where κ denotes the amount of the warning effect (in ms), this becomes

$$E[RT_{crossmodal}] = E[S_1] + E[S_2|I^c \cap W^c] - P(I) \cdot \Delta - P(W) \cdot \kappa. \quad (15)$$

In the unimodal condition, neither integration nor warning are possible. Thus,

$$E[RT_{unimodal}] = E[S_1] + E[S_2|I^c \cap W^c], \quad (16)$$

and we arrive at a simple expression for the combined effect of multisensory integration and warning, crossmodal interaction (CI),

$$CI \equiv E[RT_{unimodal}] - E[RT_{crossmodal}] = P(I) \cdot \Delta + P(W) \cdot \kappa. \quad (17)$$

Recall that the basic assumptions of TWIN imply that for a given spatial configuration and nontarget modality there are no sign reversals or changes in magnitude of Δ across all SOA values. The same holds for κ . Note, however, that Δ and κ can separately take on positive or negative values (or zero) depending on whether multisensory integration and warning have a facilitative or inhibitory effect. Furthermore, like for the probability of integration $P(I)$, the probability of warning $P(W)$ does change with SOA.

5.1 TWIN Predictions for FAP with warning

The occurrence of a warning effect depends on intramodal characteristics of the target and the nontarget, such as modality or intensity. Assuming that increasing stimulus intensity goes along with decreased reaction time (for auditory stimuli see, e.g., Frens et al., 1995; Arndt and Colonius, 2003; for tactile, Diederich and Colonius, 2004b), TWIN makes specific predictions regarding the effect of nontarget intensity variation.

Intensity effects An intense (auditory) nontarget may have a higher chance to win the race with a headstart compared to a weak nontarget. In general, increasing the intensity of the nontarget (1) increases the probability of it functioning as a warning signal, and (2) makes it more likely for the nontarget to win the peripheral race against the target process.

SOA effects The probability of warning $P(W)$ decreases monotonically with SOA: the later the nontarget is presented the smaller are its chances to win the race against the target with some headstart γ . This differs from the non-monotonic relationship predicted between $P(I_{FAP})$ and SOA (see above). It is interesting to note that the difference in how $P(I)$ and $P(W)$ should depend on SOA is, in principle, empirically testable without any distributional assumptions by manipulating the conditions of the experiment. Specifically, if target and nontarget are presented in two distinct spatial conditions, ipsilateral and contralateral, say, one would expect Δ to take on two different values, Δ_i and Δ_c , whereas $P(W) \cdot \kappa$, the expected non-spatial warning effect, should remain the same under both conditions. Subtracting the corresponding crossmodal interaction terms then gives, after canceling the warning effect terms (Eq. 17),

$$CI_i - CI_c = P(I) \cdot (\Delta_i - \Delta_c). \quad (18)$$

This expression is an observable function of SOA and, since the factor $\Delta_i - \Delta_c$ does not depend on SOA by Assumption B3, it should exhibit the same functional form as $P(I)$: increasing and then decreasing (see Fig. 1, medium left panel).

Context effects The magnitude of the warning effect may be influenced by the experimental design. Specifically, presenting nontargets from different modalities in two distinct presentation modes, e.g., blocking or mixing the modality of the auditory and tactile nontar-

gets within an experimental block of trials, such that supposedly no changes in the expected amount of multisensory integration should occur, then subtraction of the corresponding CI values yields, after canceling the integration effect terms,

$$CI_{blocked} - CI_{mixed} = P(W) \cdot (\kappa_{mixed} - \kappa_{blocked}), \quad (19)$$

a quantity that should decrease monotonically with SOA because $P(W)$ does.

The extension of the model to include warning effects has been probed for both auditory and tactile nontargets. Concerning the warning assumptions, no clear superiority of Version A over Version B was found in the data. For detailed results on all of the tests described above, we refer to Diederich and Colonius, (2008a).

SOA and Intensity – Quantitative Predictions To illustrate the predictions of TWIN with warning for mean SRT we choose the following set of parameters. As before, the intensity parameter for the visual modality is set to $1/\lambda_V = 50$ [ms] and to $1/\lambda_A = 10, 30, 70,$ or 90 [ms] for the (auditory) nontarget, the parameter for second stage processing time when no integration and no warning occurs, $\mu \equiv E[S_2|I^c \cap W^c]$, is set to 100 ms, and the time window of integration to 200 ms. The parameter for multisensory integration is set to $\Delta_i = 20$ ms for bimodal stimuli presented ipsilaterally, and κ is set to 5 ms.

6 Conclusion: Open questions and future directions

The main contribution of the TWIN framework so far is to provide an estimate of the multisensory integration effect - and, for the extended model, also of a possible warning effect - that is "contaminated" neither by a specific stimulus onset asynchrony nor by intramodal stimulus properties like intensity. This is achieved through factorizing⁴ expected crossmodal interaction (ECI) into the probability of interaction in a given trial $P(I)$ times the amount of interaction Δ (cf. Eq. 2), the latter being measured in ms. Some potential

⁴Strictly speaking, this only holds for the focused attention version of TWIN; for the redundant target version, an estimate of the amount of statistical facilitation is required and can be attained empirically (cf. Colonius and Diederich, 2006).

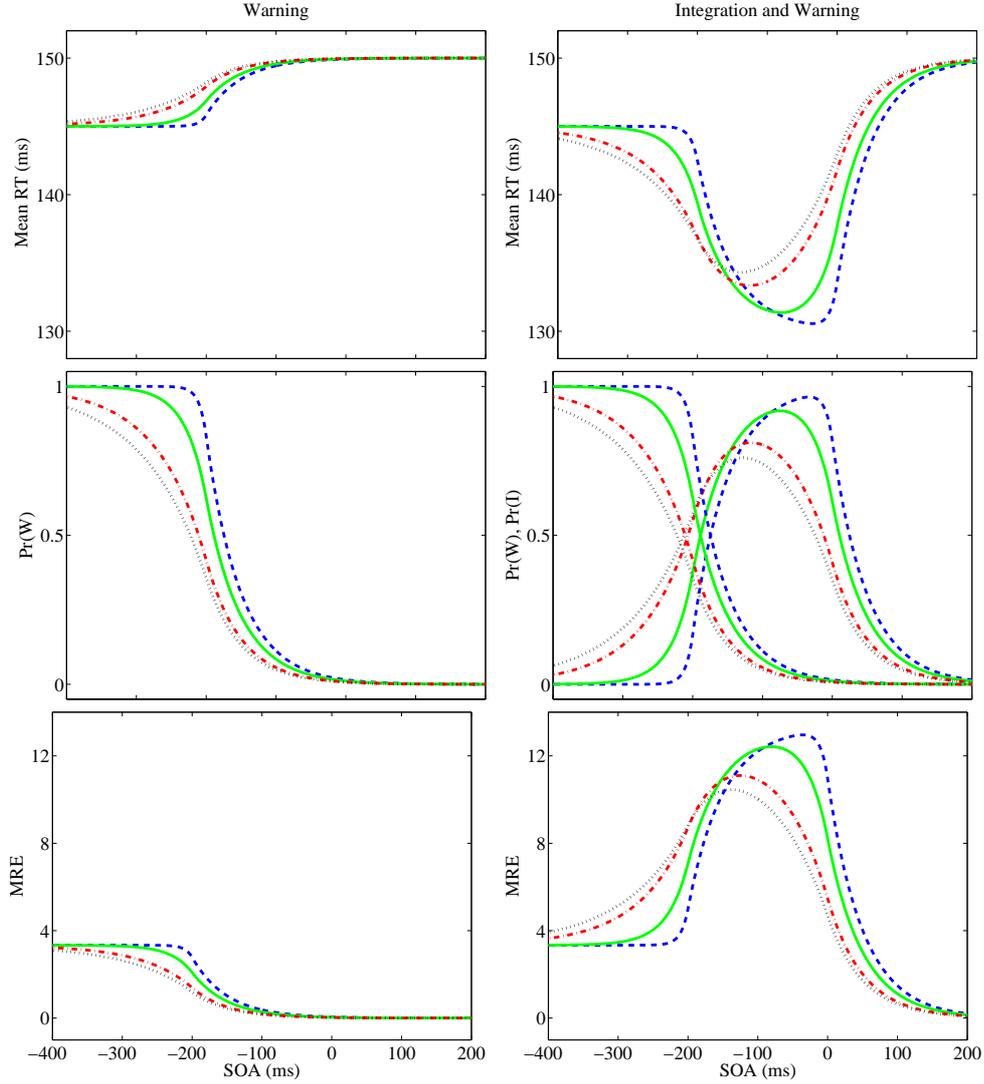


Figure 3: TWIN predictions for focused attention paradigm (FAP) when only warning occurs (left panels) and when both integration and warning occur (right panels). Parameters are chosen as before: $1/\lambda_V = 50$ and $\mu = 100$, resulting in a mean RT for the visual stimulus of 150 ms. The peripheral processing times for the auditory stimuli are $1/\lambda_A = 10$ (blue dashed line); $1/\lambda_A = 30$ (green solid); $1/\lambda_A = 70$ (red dash-dotted); and $1/\lambda_A = 90$ (black dotted).

extensions of the TWIN framework are discussed next.

While the functional dependence of $P(I)$ on SOA and stimulus parameters is made explicit in the rules governing the opening and closing of the time window, the TWIN model framework as such does not stipulate a mechanism for determining the actual amount of in-

teraction. By assumption (B4), Δ depends on crossmodal features like, for example, spatial distance between the stimuli of different modalities, and by systematically varying the spatial configuration some insight into the functional dependence can be gained (e.g., Diederich and Colonius, 2007b). Given the diversity of intersensory interaction effects, however, it would be presumptuous to aim at a single universal mechanism for predicting the amount of Δ . This does not preclude incorporating multisensory integration mechanisms into the TWIN framework within a specific context like a spatial orienting task. Such an approach, which includes stipulating distributional properties of second stage processing time in a given situation, would bring along the possibility of a stronger quantitative model test, namely at the level of the entire observable reaction time distribution rather than at the level of means only.

Another direction is to extend the TWIN framework to account for additional experimental paradigms. For example, in many studies a subject's task is not simply to detect the target but to perform a speeded discrimination task between two stimuli (Driver and Spence, 2004). Modeling this task implies not only a prediction of reaction time but also of the frequency of a correct or incorrect discrimination response. Traditionally, such data have been accommodated by assuming an evidence accumulation mechanism sequentially sampling information from the stimulus display either favoring response option A or B, say, and stopping as soon as a criterion threshold for one or the other alternative has been reached. A popular subclass of these models are the *diffusion models* which have been considered as models of multisensory integration early on (Diederich, 1995; 2008). At this point, however, it is an open question how this approach can be reconciled with the TWIN framework.

One of the most intriguing neurophysiological findings has been the suppression of multisensory integration ability of superior colliculus (SC) neurons by a temporary suspension of corticotectal inputs from the anterior ectosylvian sulcus (AES) and the lateral suprasylvian sulcus (rLS) (Clemo and Stein, 1986; Jiang et al., 2001). A concomitant effect on multisensory orientation behavior observed in cat (Jiang et al., 2002) suggests the existence of more general cortical influences on multisensory integration. Up to now, there is no explicit provision of a top-down mechanism in the TWIN framework. Note, however, that the influence of task instruction (focused attention vs. redundant target paradigm) is implicitly incorpo-

rated in TWIN since the probability of integration is supposed to be computed differently under otherwise identical stimulus conditions (cf. Section 4.4). It is a challenge for future development to demonstrate that the explicit incorporation of top-down processes can be reconciled with the two-stage structure of the TWIN framework.

7 References

- Amlôt, R., Walker, R., Driver, J., & Spence, C. (2003). Multimodal visual-somatosensory integration in saccade generation. *Neuropsychologia*, *41*, 1–15.
- Anastasio, T. J., Patton, P. E., & Belkacem-Boussaid, K. (2000). Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Computation*, *12*, 1165–1187.
- Arndt A, Colonius H (2003). Two separate stages in crossmodal saccadic integration: Evidence from varying intensity of an auditory accessory stimulus. *Exp Brain Res* 150:417–426
- Bell AH, Meredith A, Van Opstal AJ, Munoz DP (2005) Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *J Neurophysiol* 93: 3659–3673
- Clemo HR, Stein BE (1986) Effects of cooling somatosensory corticotectal influences in cat. *J Neurophys* 55:1352-1368
- Colonius H, Arndt P (2001) A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys*, *63*: 126–147.
- Colonius H, Diederich A (2002) A maximum-likelihood approach to modeling multisensory enhancement. In: T.G. Ditterich, S.Becker, & Z. Ghahramani, *Advances in Neural Information Processing Systems 14*, Cambridge: MIT Press.
- Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *J Cogn Neurosci* 16:1000–1009
- Colonius H, Diederich A (2006) Race model inequality: Interpreting a geometric measure of the amount of violation. *Psychol Review* 113 (1):148–54
- Colonius H, Diederich A, Steenken R (2009) Time-window-of-integration (TWIN) model for saccadic reaction time: Effect of auditory masker level on visual-auditory spatial interaction in elevation. *Brain Topography (Special Issue on Multisensory Integration)*

- Colonius, H., Vorberg, D. (1994) Distribution inequalities for parallel models with unlimited capacity. *Journal of Mathematical Psychology*, 38, 35–58
- Corneil BD, Munoz DP (1996) The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci* 16:8193–8207
- Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J Neurophysiol* 88:438–454
- Diederich A (1995) Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *J Math Psychol*, 39:197–215
- Diederich A (2008) A further test on sequential sampling models accounting for payoff effects on response bias in perceptual decision tasks. *Percept Psychophys*, 70(2):229–256
- Diederich A, Colonius H (2004a) Modeling the time course of multisensory interaction in manual and saccadic responses. In: G Calvert, C Spence, BE Stein (eds) *Handbook of Multisensory Processes*. MIT Press, Cambridge, MA
- Diederich A, Colonius H (2004b) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Percept Psychophys* 66(8):1388–404
- Diederich A, Colonius H (2007a) Why two "distractors" are better than one: modeling the effect of nontarget auditory and tactile stimuli on visual saccadic reaction time. *Exp Brain Res* 179:43–54
- Diederich A, Colonius H (2007b) Modeling spatial effects in visual-tactile saccadic reaction time. *Percept Psychophys* 69(1):56–67
- Diederich A, Colonius H (2008a) Crossmodal interaction in saccadic reaction time: Separating multisensory from warning effects in the time window of integration model. *Exp Brain Res* 186:1-22
- Diederich A, Colonius H (2008b) Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, 46:2556–2562.
- Diederich A, Colonius H, Bockhorst D, Tabeling S (2003) Visual-tactile spatial interaction in saccade generation. *Exp Brain Res* 148: 328–337

- Doyle MC, Walker R (2002) Multisensory interactions in saccade target selection: Curved saccade trajectories *Exp Brain Res* 142:116–130
- Driver J, Spence C (2004) Crossmodal spatial attention: Evidence from human performance. In C Spence and J Driver (Eds), *Crossmodal space and crossmodal attention* (pp. 179–220). Oxford: Oxford University Press.
- Eimer M (2001) Crossmodal links in spatial attention between vision, audition, and touch: Evidence from event-related brain potentials. *Neuropsychologia* 39:1292–1303
- Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57:802–816
- Harrington LK, Peck CK (1998) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res* 122:247–252
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol*, 63:289-293.
- Hughes HC, Nelson MD, Aronchick DM (1998) Spatial characteristics of visual-auditory summation in human saccades. *Vis Res* 38:3955–3963
- Hughes, H. C., Reuter-Lorenz, P.-A., Nozawa, G., Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 131–153.
- Jiang W, Wallace MT, Jiang H, Vaughan JW, Stein BE (2001) Two cortical areas mediate multisensory integration in superior colliculus neurons. *J Neurophys* 85:506–522
- Jiang W, Jiang H, Stein BE (2002). Two cortical areas facilitate multisensory orientation behaviour. *J Cog Neurosc* 14:1240–1255
- Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, et al (2007) Causal Inference in Multisensory Perception. *PLoS ONE* 2(9): e943. doi:10.1371/journal.pone.0000943
- Klein R, Kingstone A (1993) Why do visual offsets reduce saccadic latencies? *Behav Brain Sci* 16(3):583584

- Luce RD (1986) *Response Times: Their Role in Inferring Elementary Mental Organization*, New York: Oxford University Press.
- McGurk H, Mac Donald J (1976) Hearing lips and reading voices. *Nature*, 264:746-748.
- Meredith MA (2002) On the neural basis for multisensory convergence: a brief overview. *Cogn. Brain Res* 14:31–40
- Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J of Neurophys*, 56:640-662.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.*, 10:3215-3229.
- Miller JO (1982) Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, 14:247-279.
- Munoz DP, Wurtz RH (1995). Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *J. Neurophysiol.*, 73:2313–2333.
- Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cogn Brain Res* 25: 499–507
- Nickerson RS (1973) Intersensory facilitation of reaction time: energy summation or preparation enhancement. *Psychol Rev*, 80:489-509.
- Raab DH (1962) Statistical facilitation of simple reaction times. *Trans NY Acad Sci* 24:574–590
- Reuter-Lorenz PA, Hughes HC, Fendrich R (1991) The reduction of saccadic latency by prior offset of the fixation point: an analysis of the gap effect. *Percept Psychophys* 49(2):167–175
- Ross SM, Ross LE (1981) Saccade latency and warning signals: effects of auditory and visual stimulus onset and offset. *Percept Psychophys* 29(5):429-437
- Rowland BA, Stein BE (2008) Temporal profiles of response enhancement in multisensory integration. *Frontiers in Neuroscience* 2:218–224

- Schweickert R , Fisher DL, Sung K (in press) Discovering Cognitive Architecture by Selectively Influencing Mental Processes. London: World Scientific Publishing
- Sinclair C, Hammond GR (2009) Excitatory and inhibitory processes in primary motor cortex during the foreperiod of a warned reaction time task are unrelated to response expectancy. *Exp Brain Res* 194:103–113
- Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. *Curr Biol* 13:R519–R521
- Stein BE, Meredith MA (1993) The merging of the senses. The MIT Press, Cambridge MA
- Stein BE, Huneycutt WS, Meredith MA (1988). Neurons and behavior: The same rules of multisensory integration apply. *Brain Res* 448:355–358
- Stein BE, Jiang W, Stanford TR (2004). Multisensory integration in single neurons in the midbrain. In: G Calvert, C Spence, BE Stein (eds) *Handbook of Multisensory Processes*. MIT Press, Cambridge, MA, 243-264
- Sternberg S (2001) Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, 106:147–246
- Todd, JW (1912) Reaction to multiple stimuli, in *Archives of Psychology, No. 25. Columbia Contributions to Philosophy and Psychology*, (R. S. Woodworth, ed.), Vol. XXI No. 8, New York: The Science Press.
- Townsend, J. T., Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39, 321-359.
- Van Opstal AJ, Munoz DP (2004) Auditory-visual interactions subserving primate gaze orienting. In: G Calvert, C Spence, BE Stein (eds.) *Handbook of Multisensory Processes*, Cambridge, MA, MIT Press (pp. 373–393)
- Van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598–607
- Van Zandt T (2002) Analysis of response time distributions. In: H Pashler (Ed) *Stevens' Handbook of Experimental Psychology, Volume 4, 3rd Edition*, Wiley & Son, Inc.

Whitchurch EA, Takahashi TT (2006) Combined auditory and visual stimuli facilitate head saccades in the barn owl (*Tyto alba*). *J Neurophysiol* 96:730–745

8 Appendix

8.1 Deriving the probability of interaction in TWIN

The peripheral processing times V for the visual and A for the auditory stimulus have an exponential distribution with parameters λ_V and λ_A , respectively. That is,

$$\begin{aligned} f_V(t) &= \lambda_V e^{-\lambda_V t}, \\ f_A(t) &= \lambda_A e^{-\lambda_A t} \end{aligned}$$

for $t \geq 0$, and $f_V(t) = f_A(t) \equiv 0$ for $t < 0$. The corresponding distribution functions are referred to by $F_V(t)$ and $F_A(t)$.

8.1.1 Focused Attention Paradigm

The visual stimulus is the target and the auditory stimulus is the nontarget. By definition,

$$\begin{aligned} P(I_{FAP}) &= Pr(A + \tau < V < A + \tau + \omega) \\ &= \int_0^\infty f_A(x) \{F_V(x + \tau + \omega) - F_V(x + \tau)\} dx, \end{aligned}$$

where τ denotes the SOA value and ω is the width of the integration window. Computing the integral expression requires that we distinguish between three cases for the sign of $\tau + \omega$:

(i) $\tau < \tau + \omega < 0$

$$\begin{aligned} P(I_{FAP}) &= \int_{-\tau-\omega}^{-\tau} \lambda_A e^{-\lambda_A x} \{1 - e^{-\lambda_V(x+\tau+\omega)}\} dx \\ &\quad + \int_{-\tau}^\infty \lambda_A e^{-\lambda_A x} \{e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)}\} dx \\ &= \frac{\lambda_V}{\lambda_V + \lambda_A} e^{\lambda_A \tau} (-1 + e^{\lambda_A \omega}); \end{aligned}$$

(ii) $\tau < 0 < \tau + \omega$

$$\begin{aligned}
P(I_{FAP}) &= \int_0^{-\tau} \lambda_A e^{-\lambda_A x} \{1 - e^{-\lambda_V(x+\tau+\omega)}\} dx \\
&\quad + \int_{-\tau}^{\infty} \lambda_A e^{-\lambda_A x} \{e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)}\} dx \\
&= \frac{1}{\lambda_V + \lambda_A} \{ \lambda_A(1 - e^{-\lambda_V(\omega+\tau)}) + \lambda_V(1 - e^{-\lambda_A \tau}) \};
\end{aligned}$$

(iii) $0 < \tau < \tau + \omega$

$$\begin{aligned}
P(I_{FAP}) &= \int_0^{\infty} \lambda_A e^{-\lambda_A x} \{e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)}\} dx \\
&= \frac{\lambda_A}{\lambda_V + \lambda_A} \{e^{-\lambda_V \tau} - e^{-\lambda_V(\omega+\tau)}\}
\end{aligned}$$

The mean RT for crossmodal stimuli is

$$\begin{aligned}
E[RT_{V,A,\tau}] &= E[V] + E[S_2 | I_{FAP}^c] - P(I_{FAP}) \cdot \Delta \\
&= \frac{1}{\lambda_V} + \mu - P(I_{FAP}) \cdot \Delta
\end{aligned}$$

and the mean RT for the visual target is

$$E[RT_V] = \frac{1}{\lambda_V} + \mu,$$

where $1/\lambda_V$, the mean of the exponential distribution, is the mean RT of the first stage and μ is the mean RT of the second stage when no interaction occurs.

8.1.2 Redundant Target Paradigm

The visual stimulus is presented first and the auditory stimulus second. By definition,

$$P(I_{RTP}) = Pr\{\max(V, A + \tau) < \min(V, A + \tau) + \omega\}$$

If the visual stimulus wins:

$$(i) \quad 0 \leq \tau \leq \omega$$

$$\begin{aligned} P(I_{RTP_V}) &= \int_0^\tau \lambda_V e^{-\lambda_V x} (1 - e^{-\lambda_A(x+\omega-\tau)}) dx \\ &\quad + \int_\tau^\infty \lambda_V e^{-\lambda_V x} (1 - e^{-\lambda_A(x+\omega-\tau)} - (1 - e^{-\lambda_A(x-\tau)})) dx \\ &= \frac{1}{\lambda_V + \lambda_A} \{ \lambda_V (1 - e^{\lambda_A(-\omega+\tau)}) + \lambda_A (1 - e^{(-\lambda_V \tau)}) \}; \end{aligned}$$

$$(ii) \quad 0 < \omega \leq \tau$$

$$\begin{aligned} P(I_{RTP_V}) &= \int_{\tau-\omega}^\tau \lambda_V x (1 - e^{-\lambda_A(x+\omega-\tau)}) dx \\ &\quad + \int_\tau^\infty \lambda_V e^{-\lambda_V x} (1 - e^{-\lambda_A(x+\omega-\tau)} - (1 - e^{-\lambda_A(x-\tau)})) dx \\ &= \frac{\lambda_A}{\lambda_V + \lambda_A} \{ e^{-\lambda_V \tau} \cdot (-1 + e^{\lambda_V \omega}) \} \end{aligned}$$

If the auditory stimulus wins: $0 < \tau \leq \tau + \omega$ and

$$P(I_{RTP_A}) = \int_0^\infty \lambda_A e^{-\lambda_A x} \{ e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)} \} dx$$

$$= \frac{\lambda_A}{\lambda_V + \lambda_A} \{e^{-\lambda_V \tau} - e^{-\lambda_V(\omega+\tau)}\}$$

The probability that the visual or the auditory stimulus wins is therefore

$$P(I_{RTP}) = P(I_{RTP_V}) + P(I_{RTP_A}).$$

The mean RT for crossmodal stimuli is

$$\begin{aligned} E[RT_{VA,\tau}] &= E[\min(V, A + \tau)] + E[S_2 | I_{RTP}^c] - P(I_{RTP}) \cdot \Delta \\ &= \frac{1}{\lambda_V} - e^{-\lambda_V \tau} \cdot \left(\frac{1}{\lambda_V} - \frac{1}{\lambda_V + \lambda_A} \right) + \mu - P(I_{RTP}) \cdot \Delta \end{aligned}$$

and the mean RT for the visual and auditory stimulus is

$$E[RT_V] = \frac{1}{\lambda_V} + \mu,$$

and

$$E[RT_A] = \frac{1}{\lambda_A} + \mu,$$

8.1.3 Focused Attention and Warning

By definition,

$$\begin{aligned} P(W) &= Pr(A + \tau + \gamma_A < V) \\ &= \int_0^\infty f_A(x) \{1 - F_V(x + \tau + \gamma_A)\} dx \\ &= 1 - \int_0^\infty f_A(x) F_V(a + \tau + \gamma_A) dx. \end{aligned}$$

Again, we need to consider different cases:

(i) $\tau + \gamma_A < 0$

$$P(W) = 1 - \int_{-\tau-\gamma_A}^\infty \lambda_A e^{-\lambda_A a} \{1 - e^{-\lambda_V(a+\tau+\gamma_A)}\} da$$

$$= 1 - \frac{\lambda_V}{\lambda_V + \lambda_A} e^{\lambda_A(\tau + \gamma_A)};$$

(ii) $\tau + \gamma_A \geq 0$

$$\begin{aligned} P(W) &= 1 - \int_0^\infty \lambda_A e^{-\lambda_A a} \{1 - e^{-\lambda_V(a + \tau + \gamma_A)}\} da \\ &= \frac{\lambda_A}{\lambda_V + \lambda_A} e^{-\lambda_V(\tau + \gamma_A)}. \end{aligned}$$

The mean RT for crossmodal stimuli is

$$\begin{aligned} E[RT_{VA,\tau}] &= E[V] + E[S_2 | I_{FAP}^c] - P(I_{FAP}) \cdot \Delta - P(W) \cdot \kappa \\ &= \frac{1}{\lambda_V} + \mu - P(I_{FAP}) \cdot \Delta - P(W) \cdot \kappa \end{aligned}$$

where $1/\lambda_V$ is the mean RT of the first stage, μ is the mean RT of the second stage when no interaction occurs, $P(I_{FAP}) \cdot \Delta$ is the expected amount of intersensory interaction, and $P(W) \cdot \kappa$ is the expected amount of warning.