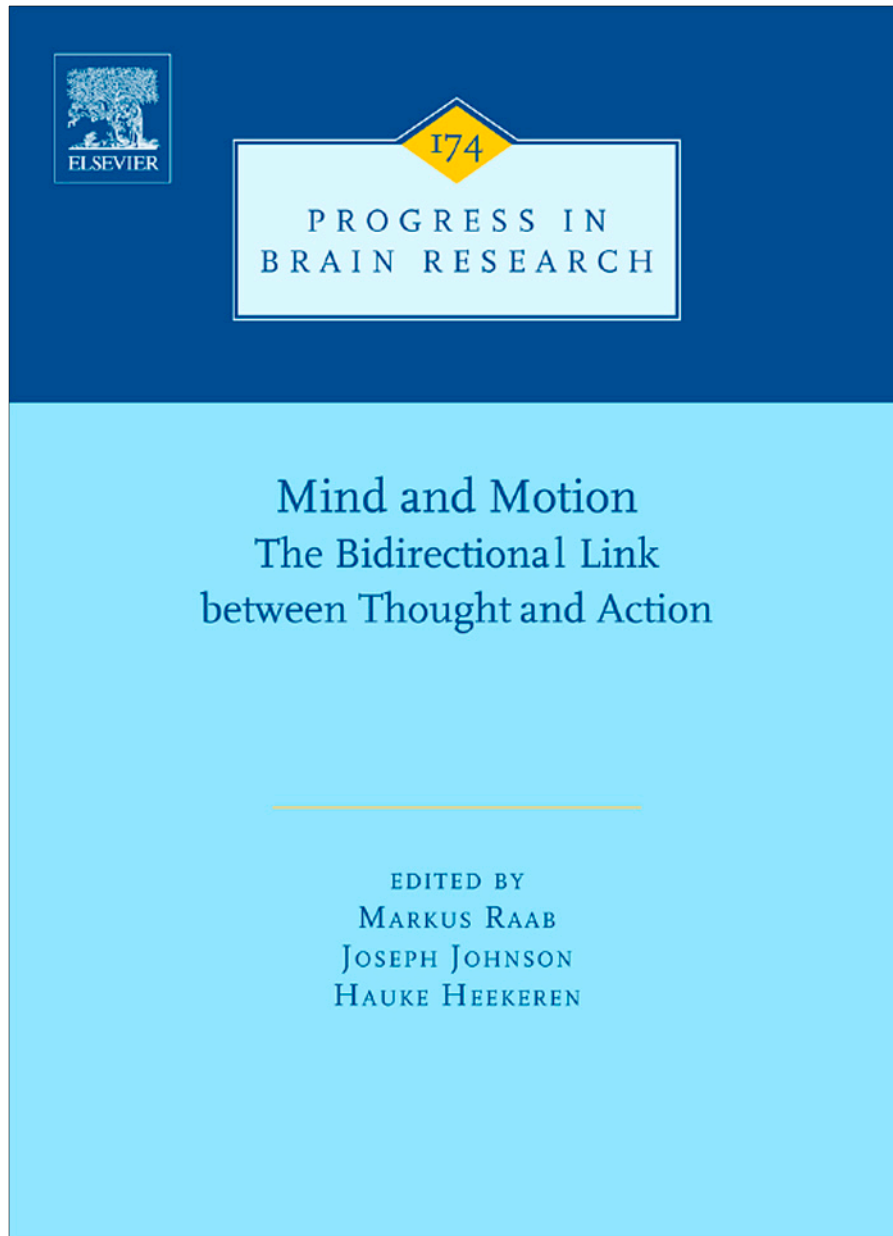


**Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.**

This chapter was originally published in the book *Progress in Brain Research*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>

From Adele Diederich and Hans Colonius, Crossmodal interaction in speeded responses: time window of integration model. In: Markus Raab, Joseph G. Johnson and Hauke R. Heekeren, editors: *Progress in Brain Research*, Vol 174, *Mind and Motion: The Bidirectional Link between Thought and Action*, Markus Raab, Joseph G. Johnson and Hauke R. Heekeren. The Netherlands: Elsevier, 2009, pp. 119–135.

ISBN: 978-0-444-53356-2

© Copyright 2009 Elsevier BV.

Elsevier

CHAPTER 11

Crossmodal interaction in speeded responses: time window of integration model

Adele Diederich^{1,*} and Hans Colonius²

¹*School of Humanities and Social Sciences, Jacobs University, Bremen, Germany*
²*Department of Psychology, Oldenburg University, Oldenburg, Germany*

Abstract: Saccadic reaction time (SRT) to a visual stimulus tends to be faster when an auditory and/or somatosensory stimulus is presented in close temporal or spatial proximity, even when participants are instructed to ignore the accessory input (focused attention task). The time course of SRT as a function of stimulus onset asynchrony (SOA) is consistent with the time-window-of-integration (TWIN) model assuming a peripheral stage of parallel processing in separate sensory channels followed by a secondary stage of multisensory integration. TWIN has been shown to account for effects of the spatial configuration of the stimuli, for the effect of increasing the number of nontargets presented together with the target, for a possible warning effect of the nontarget, for effects of increasing the intensity of the nontarget, and for the effect of background noise on multisensory integration. Moreover, it has been able to accommodate some effects of aging on multisensory integration.

There is empirical support for TWIN's tenet of the separability between spatial and temporal factors on multisensory integration. Besides presenting many features of TWIN within the context of crossmodal interaction modeling efforts, some possible directions on how the TWIN framework could serve to elucidate the link between perception and action are shown.

Keywords: multisensory integration; warning effect; time-window-of-integration; response time; saccadic eye movement

Introduction

Adaptive behavior depends on the ability of the perceptual system to deliver information about ongoing events in the environment rapidly. This information typically arrives via different sensory channels and has to be integrated to produce a coherent internal representation of the outside

world. Our daily experience provides many examples for successful multisensory integration. We identify objects by visual information such as shape, color, or size but also by auditory or tactile information, for example a singing bird, a fire engine, velvet. A simple act like drinking sweet hot green tea from a Styrofoam cup involves integrating information from various modalities such as vision, taste, olfaction, and touch. Sensory information arriving from various modalities needs to be processed in parallel for adequate action.

Action execution also depends on crossmodal integration. For instance, kicking a ball requires

*Corresponding author.
Tel.: +49-421-200-3431; Fax: +49-421-200-3303;
E-mail: a.diederich@jacobs-university.de

the integration of visual, proprioceptive, and tactile information, and performance may critically depend on how well this information is integrated across the different reference frames. In keeping with the ball game example, motion of the body in physical space, or even a simple movement of the head, will in turn modify the incoming crossmodal stream of information requiring yet another crossmodal integration process. Obviously, action execution and crossmodal integration are closely intertwined components of a bidirectional dynamic process involving in an organism adapting to the environment.

Studying multisensory integration is an emerging field in neuroscience strongly influencing behavioral scientists traditionally investigating crossmodal interaction. Neurophysiologists are predominantly interested in where and how different sensory modalities are integrated in the brain. For example, they found multisensory neurons in various animals, for example, in the superior colliculus (SC) of the cat. The neurons in the deep layers of the superior colliculus (DLSC) are important in the control of eye movements and, in general, for directing exogenous attention. In fact, a large majority of multisensory neurons in cat DLSC show an enhanced response to particular combinations of visual, auditory, and tactile stimuli relative to the best modality-specific responses. For instance, Meredith and Stein (1986) found that a visual stimulus evoked responses on only six of 16 trials in a bimodal neuron and these responses were composed of few impulses. An auditory stimulus also evoked weak and unreliable responses. However, their combination produced a vigorous response on every trial. The mean number of impulses was increased by up to 1207% compared to the most effective unimodal stimulus. Information about stimulus location is represented topographically within the structure by an orderly arrangement of neurons according to the location of their respective receptive fields (RFs). The spatial register among the different sensory maps is formed by the multisensory neurons whose different RFs are in register with one another (for a review see Stein and Meredith, 1993). In addition, the SC contains a motor map composed

of output neurons coding appropriate eye movements (Sparks, 1986), that is, the locus of activity in the motor map encodes a movement command that reorients the eyes (and the head) a given distance in a particular direction. Thus, if the goal is to translate a sensory target into an appropriate motor command, the alignment of the visual, auditory, and tactile maps to each other and to the motor map is critical.

In the behavioral sciences and, in particular, in the psychological literature, numerous effects of crossmodal stimulation on perception and action have been described over more than a century (cf. Welch and Warren, 1986). Prominent examples are the McGurk effect (McGurk and MacDonald, 1976), the ventriloquism effect (Howard and Templeton, 1966; for the effect in humans and monkeys, e.g., Wood and Recanzone, 2004), synesthesia, and many more (see Calvert et al., 2004, for an overview). While many phenomena, like synesthetic experiences, can only be described at a qualitative level, other crossmodal effects are amenable to standard psychophysical measurement techniques.

One such measure is response time (RT). It is a ubiquitous measure used in experimental psychology for more than 150 years to investigate hypotheses about the mental and motor processes leading to the generation of a response. Thus, it comes as no surprise that some of the first psychological studies of intersensory interaction employed RT as a measure to assess the effect of combining stimuli from different modalities and of varying their intensities (Todd, 1912). RT analysis becomes most powerful in uncovering underlying processes when it is based on explicit quantitative hypotheses (Luce, 1986). One simple reason for the preponderance of response time as a measure of crossmodal effects is that, whatever happens within an organism processing stimulus information from several sensory modalities, it must unfold over time. One of the main experimental findings in behavioral studies is that the occurrence of crossmodal effects critically depends on the temporal arrangement of the stimulus sets. For example, the speed-up of the RT to a visual stimulus resulting from presenting, say, an accessory auditory stimulus typically becomes greatest

when the visual stimulus precedes the auditory by an interval that equals the difference in RT between response to the visual stimulus alone and the auditory stimulus alone (Hershenson, 1962). For instance, assuming a mean RT to a visual stimulus of 160 ms and to an auditory of 120 ms, the largest amount of facilitation is expected when the visual stimulus precedes the auditory by $160\text{ ms} - 120\text{ ms} = 40\text{ ms}$. Interestingly, temporal integration rules similar to this one have been discerned for the responses at the level of individual multisensory neurons as well. Specifically, bimodal cells in cat superior colliculus (SC) show maximal levels of response enhancement when the peak discharge periods evoked by each modality overlap in time (Meredith et al., 1987). This overlap is often correlated not with simultaneous stimulus presentation but with differences in stimulus presentation that match their latency differences. Since multisensory SC neurons project to premotor areas of the brainstem that control orientation of the eyes, pinnae, and head, the temporal relationships of stimulus complexes that control the activity of these neurons should ultimately contribute to orientation behavior as well. Of course, individual neurons differ in their temporal response properties, and behavior is the result of a large number of neurons at different sites of the brain. Nonetheless, it is a distinct possibility that the temporal integration rules observed in behavioral experiments can ultimately be deduced from certain principles holding at the level of the individual neuron or of an assembly of neurons.

In the following, we describe mathematical models of RT that have been developed to explain and to predict crossmodal stimulus effects resulting from a manipulation of their spatial-temporal configuration and of physical stimulus dimensions like intensity. One feature of the models is that they are based on a few, relatively simple principles — if not mathematically, then at least conceptually — so that they certainly do not reflect the full complexity of the underlying multisensory processes. Note that this should be considered a virtue rather than a defect at this stage: these models can be tested rigorously and, hopefully, converge on those principles that

constitute a valid framework for a more detailed account to be developed as more evidence is collected.

Redundant target versus focused attention paradigm

Two different experimental paradigms have been utilized to measure RT to a crossmodal stimulus set. In the *redundant target* (also known as *divided-attention*) paradigm, stimuli from different modalities (e.g., a tone and a light) 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 (e.g., a tone plus a light) is faster than in either of the unimodal conditions (e.g., a tone or a light). In the *focused attention* paradigm, 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 (e.g., such as the visual), and to ignore the remaining nontarget stimulus (e.g., the tactile or the auditory). In the latter setting, when a stimulus of a nontarget modality (e.g., a tone), 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 response to the target stimulus: depending on the exact spatiotemporal configuration of target and nontarget, the effect can be a speed-up or an inhibition of RT (for an overview see, for example, Diederich and Colonius, 2004a, b).

Note that the distinction between the redundant target and the focused attention paradigm is not only an interesting experimental variation as such, but it may also provide an important theoretical perspective. In fact, since the stimuli can be chosen to be physically identical in both paradigms, any differences observed in the corresponding responses would have to be due to the instructions only, thereby giving the researcher an opportunity to separate top-down from bottom-up processes in the multisensory integration mechanism.

Separate activation versus coactivation models

Usually, observable RT is divided into two additive components, that is,

$$RT = S + R \quad (1)$$

where the random variable S denotes stimulus processing time proper, and R is base time random variable denoting response preparation, motor delay, etc. Here, RT will mostly be considered at the level of expectation only, instead of the entire probability distribution. This simplifies the discussion significantly; in particular, the expected value of R , $E[R]$, can be treated as a constant which, when added, often does not modify the main results of interest and which will, therefore, often not be mentioned explicitly. Numerous models have been suggested to account for the crossmodal effects in RTs. Two different approaches are commonly distinguished, *separate activation* versus *coactivation* models. Separate activation models, also known as *race models*, assume (1) that presenting a crossmodal stimulus produces parallel, separate activation in different sensory channels that build to the level at which they can produce a response, and (2) that the response is triggered by the signal that reaches that level first. Assuming statistical variability in the channel processing times, separate activation models predict faster average RT to crossmodal stimuli than to unimodal stimuli because the average of the winner's processing time is smaller than the average processing time in each single channel ("statistical facilitation", cf. Raab, 1962).

For example, let T , V , and A denote the random processing time for a tactile, visual, and auditory stimulus, respectively. Then,

$$E[\min(T, V, A)] \leq \min(E[T], E[V], E[A]) \quad (2)$$

where $E[\]$ indicates the expected value of the random variables (*Jensen's inequality*, cf. Billingsley, 1979). Thus, statistical facilitation predicts faster mean responses to crossmodal stimuli as a statistical phenomenon without assuming an additional neural mechanism.

A more general test of separate activation models proposed by Miller (1982) has become

the standard tool to assess whether statistical facilitation suffices to explain the response speed-up of crossmodal stimulus presentations. For three modalities, the test is based on the inequality

$$P(\min\{T, V, A\} \leq t) \leq P(T \leq t) + P(V \leq t) + P(A \leq t) \quad (3)$$

with the same notation as above. This inequality puts an upper bound on the facilitation produced by trimodal stimuli. However, since it is the sum of three probabilities approaching a value of three for increasing values of t , this bound may not always be of use in testing separate activation. Interestingly, the upper bound can be sharpened in various ways (cf. Diederich, 1992). In any event, Miller's test has often been shown to be violated in crossmodal experiments with two modalities, and the amount of violation of the inequality has been interpreted as a measure of the strength of neural coactivation (e.g., see Colonius and Diederich, 2006).

As an alternative approach, *coactivation models* assume that activation raised in different sensory channels by presenting crossmodal stimuli is combined to satisfy a single criterion for response initiation. Coactivation models predict faster average RT to multiple stimuli compared to single stimuli because the combined activation reaches that criterion faster. Assume that presentation of a stimulus triggers a sequence of "events" occurring randomly over time. In a neurophysiological context, these events are typically interpreted as spikings of a neuron, for example, but the model can be formulated at a more abstract level (cf. Tuckwell, 1989). The only relevant property of the events is their time of occurrence, and all information about the stimulus is contained in the time course of the events. For example, the rate of the event sequence, that is, the mean number of events per unit time interval, is typically thought to be related to signal intensity. Let $N(t)$ denote the number of events that have occurred by time t after stimulus presentation. *Counter models* assume that $N(t)$ has some internal representation registering the number of events over the course of time.

Let us assume a counter model where separate counters exist for each modality, $N_V(t)$, $N_A(t)$, and $N_T(t)$, for a visual, auditory, and tactile stimulus, respectively. Presentation of a crossmodal stimulus triggers all counters to start registering their sequence of modality-specific events, and the counter that first reaches a preset criterion initiates the response. Obviously, under this assumption, we return to the class of separate activation (race) models, where a race between counters takes place and the winner determines the response.

Alternatively, in *superposition models* the counters activated by a crossmodal stimulus will be summed. For example, a visual–auditory stimulus would trigger a composite counter $N_2(t) \equiv N_V(t) + N_A(t)$, and a trimodal stimulus corresponds to $N_3(t) \equiv N_V(t) + N_A(t) + N_T(t)$, analogously. Intuitively, the more counters are combined in the composite counter the faster a fixed criterion number of counts, c , say, will be reached on average. Figure 1 illustrates this intuition.

To compute the distribution of the (random) *waiting time* S_c for the c th count to occur note that:

$$P(S_c \leq t) = P(N(t) \geq c)$$

The most tractable case to derive exact quantitative predictions from is the *Poisson (counting) process* where it is assumed that for each counter the times between successive events (*interarrival times*) are independent exponentially distributed random variables. Each Poisson process is characterized by a single constant, the *intensity parameter* λ . The expected waiting time for the c th count is then simply c/λ . Superposition models represent RT by the waiting time S_c for the c th count. The criterion c is a (bias) parameter describing the subject's strategic behavior. Specifically, requiring high accuracy from the subject,

for example, avoiding anticipation responses, may raise the criterion, whereas requiring high response speed may lower it. It is assumed to be a constant, however, over a given experimental condition (see Luce, 1986, for a comprehensive discussion).

A superposition model for response time in redundant target experiments with two modalities was proposed by Schwarz (1989). Diederich (1992, 1995) extended the model to deal with trimodal stimulus data. In the Poisson superposition model with λ_T , λ_V , and λ_A denoting the intensity parameters of the tactile, visual, and auditory stimuli, respectively, the expected waiting time for the c th count to occur when all three stimuli are presented is:

$$E[S_c | trimodal] = \frac{c}{\lambda_T + \lambda_V + \lambda_A} \quad (4)$$

At the level of mean RTs, the following model predictions are obvious:

- (1) The smaller the intensity parameter λ , the fewer counts are registered within a given time interval, and the longer it takes to reach the criterion to initiate a response. Thus, if λ is assumed to be an increasing function of stimulus intensity, mean RT to weak stimuli should be longer than to strong stimuli. For example, using a 50 dB and a 70 dB tone with $\lambda_{50} < \lambda_{70}$,

$$E[S_c | 50 \text{ dB}] = \frac{c}{\lambda_{50}} > E[S_c | 70 \text{ dB}] = \frac{c}{\lambda_{70}}$$

- (2) Mean RT to multimodal stimuli should be shorter than to unimodal stimuli, and it decreases with the number of modalities involved.

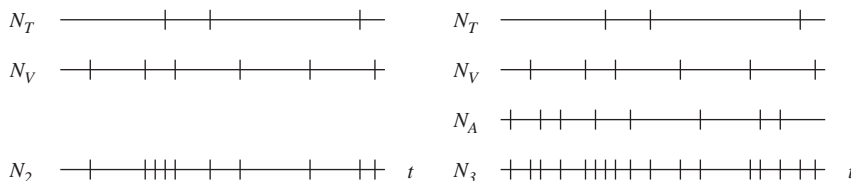


Fig. 1. Superposition of visual, auditory, and tactile counts.

Overall, at the level of the means, the model often gives a very satisfactory fit for the data. In particular, it is possible to predict mean response times in the bimodal condition by using parameter values estimated in the trimodal condition (for details see Diederich, 1992, 1995). On the negative side, the variability in the response speed, as measured by RT variance (derived in Diederich and Colonius, 1991), is not adequately captured by the model. This may be due to the fact that for the exponential distribution of the interarrival times (in the Poisson process) mean and variance are strictly coupled, and does not present evidence against the superposition model in general. Schwarz (1994) has argued that the variability in the data can be accounted for if an additional additive (motor) component, possibly with a strong negative correlation to the decision time, is introduced following the presentation in Eq. (1). Note, however, that it seems unfeasible to test such assumptions about the nonobservable bivariate distribution of (S, R) within in the RT paradigm experimentally.

The direct representation of stimulus intensity by an intensity parameter in superposition models predicts that increasing stimulus intensity should lead to ever faster responses, without being able to account for any of the typical saturation effects. Therefore, modelers have considered yet another version of the coactivation idea, the so-called *diffusion models*. The *multichannel diffusion* model resembles the superposition model in many ways. Like the latter, its mathematical foundation is a class of stochastic processes but, whereas superposition models are based on counting processes, the multichannel diffusion model is based on *Brownian motion*, or *diffusion* processes (cf. Billingsley, 1979). Generalizing the *counter* concept, response initiation depends on a stimulus-triggered activation accumulation process to cross a criterion level of activation. The level of activation varies continuously rather than in discrete counts. The main determinant of the process is its *drift rate*, μ , a parameter that can intuitively be considered as the instantaneous propensity for the activation to go up or down by an infinitesimal amount. Each presentation of a stimulus triggers the realization of a function

describing the course of activation over time, called a *trajectory*. The drift parameter is constant over time, but in a particular type of diffusion process, the *Ornstein-Uhlenbeck process* (OUP), it is a function of the activation level:

$$\mu(x) = \delta - \gamma \cdot x \quad (5)$$

where δ refers to the constant part of the drift driving the process to the criterion (absorbing boundary). Diffusion models of RT assume that δ is a monotonic function of stimulus intensity: strong stimuli have large δ values implying that the trajectories first have a tendency to be steep and to approach the criterion level quickly to initiate a response. Note, however, that for positive values of γ (the *decay parameter*) the drift $\mu(x)$ decreases faster the larger the activation level x becomes, that is, the closer activation gets to the criterion level. This is responsible for the trajectories to level off rather than to increase linearly over time. Moreover, when the stimulus signal is switched off, activation is assumed to decay to its starting level, since δ takes on a value of zero. It is assumed that activation never drops below its initial level. This decay process, which cannot be represented in a superposition/counter model has been discussed in studies of neuronal activity dynamics (Ricciardi, 1977; Tuckwell, 1989).

For crossmodal stimuli, the δ values corresponding to the unimodal stimuli are added:

$$\mu(x) = (\delta_T + \delta_V + \delta_A) - \gamma \cdot x \quad (6)$$

It is not difficult to see that the multichannel diffusion model affords the same predictions as the superposition model with respect to the effect of stimulus intensity and the number of modalities involved. Moreover, a parametric fit of the diffusion model to the bi- and trimodal data of the redundant target experiment with different SOAs was very close to the fit using the superposition model (for details see Diederich, 1992, 1995). The diffusion models seems to be more flexible in dealing with the problems arising with certain intensity variations mentioned above. First, the OUP diffusion model does not predict RT to go to zero with high enough stimulus intensity: an increase in δ can, in principle, be compensated for

by a corresponding increase in the decay parameter γ . Second, the principle of *inverse effectiveness* — whereby interaction should be largest for lower stimulus levels — could be realized by an appropriate elaboration of the drift function. For example, replacing the term $\delta_T + \delta_V + \delta_A$ in Eq. (6) by

$$(\delta_T + \delta_V + \delta_A) \left[1 + (\delta_T^{\max} - \delta_T)(\delta_V^{\max} - \delta_V) \right. \\ \left. \times (\delta_A^{\max} - \delta_A) \right] \quad (7)$$

or something similar, would yield an additive effect of intensity if at least one modality is close to maximum level, but an overadditive effect if all stimuli are far away from their maximum levels. Note that, in principle, one could develop superposition models along this line, but deriving the ensuing crossmodal mean RT predictions seems much more difficult.

The time-window-of-integration (TWIN) modeling scheme

All models considered so far are completely symmetric with respect to the stimulus modalities involved. In other words, these models have no mechanism to account for possible effects caused by distinguishing a target signal modality from an accessory, or *distractor*, signal modality in an experimental paradigm such as focused attention. The modeling scheme proposed in the following incorporates explicit assumptions about possible target/nontarget modality distinctions.

Based on neurophysiologic and behavioral findings in humans, monkeys, and cats, several authors have suggested the existence of a critical spatiotemporal “window” for multisensory integration to occur (e.g., Corneil et al., 2002; Meredith, 2002; Bell et al., 2005). Given that the strength of a multisensory effect is often lawfully related to the spatial and temporal configuration of the stimuli from different modalities (Stein and Meredith, 1993), the notion of a “window” has been proposed to be a critical determinant for multisensory integration at both neural and behavioral levels of observation. With respect to the temporal dimension, the idea simply is that the

visual and the auditory stimulus, for example, must not be presented too far away in time for bimodal integration to occur. 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 RT (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).

The initial separation of the afferent pathways for the different sensory modalities suggests that one can distinguish at least two serial stages of saccadic RT: an early, afferent stage of peripheral processing (*first stage*) followed by a compound stage of converging subprocesses (*second stage*). As shown below, in conjunction with a number of weak additional assumptions, some interesting and empirically testable predictions can be derived from this simple setup. Even under invariant experimental conditions, responses typically vary from one trial to the next due to an inherent variability of the underlying neural processes in both ascending and descending pathways. This is taken into account in the TWIN model by assuming the duration of each of the stages to be a random variable (Colonius and Diederich, 2004).

- (1) *First-stage assumption*: the *first stage* consists in a race among the peripheral neural excitations in the visual, auditory, and/or somatosensory pathways triggered by a crossmodal stimulus complex.

Since the first stage refers to very early sensory processing, random processing times for visual, auditory, and somatosensory stimuli are assumed to be *statistically independent*.

- (2) *Second-stage assumption*: the *second stage* comprises neural integration of the input and preparation of a response. Crossmodal interaction manifests itself in an increase or decrease of second-stage processing time.

Thus, the model retains the notion of a race but restricts it to the very first stage of stimulus processing. The assumption of only two stages is

certainly an oversimplification. Note, however, that the second stage is defined by default: it includes all subsequent, possibly overlapping, processes that are not part of the peripheral processes in the first stage.

- (3) *TWIN assumption*: multisensory integration occurs *only* if the peripheral processes of the first stage all terminate within a given time interval, the “window of integration.”

The two-stages of the TWIN model suggests an additional, important assumption about the effects of spatial and temporal factors:

- (4) *Assumption of spatiotemporal separability*: the amount of interaction in second-stage processing time is a function of the spatial configuration of the stimuli, but it does not depend on their (physical) presentation asynchrony (SOA).

Interaction, if it occurs at all, will be either inhibition or facilitation depending on both target and nontarget position. Typically, any facilitation decreases with the distance between the stimuli. More specific hypotheses about the effect of the spatial configuration on the amount of interaction have been studied in Diederich and Colonius (2007a, b).

The window of integration acts like a filter determining whether afferent information delivered from different sensory organs is registered close enough in time to allow for multisensory integration. Passing the filter is a necessary, but not a sufficient condition for multisensory integration to occur. The reason is that multisensory integration also depends on the spatial configuration of the stimuli. However, rather than assuming the existence of a joint spatial–temporal window of integration permitting interaction to occur only for both spatially and temporally neighbored stimuli, the TWIN model allows for multisensory integration to occur even for rather distant stimuli (of different modalities) as long as they fall within the time window. Such interaction will typically be an inhibition or only a small facilitation. This arrangement affords more flexibility in a complex environment. 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.

Quantifying multisensory integration in the TWIN Model

According to the two-stage assumption, total RT in the crossmodal condition can be written as a sum of two random variables:

$$RT_{crossmodal} = S_1 + S_2 \quad (8)$$

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

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

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

$$E[RT_{crossmodal}] = E[S_1] + E[S_2|I^c] - \Pr[I] \cdot \Delta \quad (9)$$

The term $\Pr[I] \cdot \Delta$ can be interpreted as a measure of the expected RT speed-up in the second stage with positive Δ values corresponding to facilitation, negative values to inhibition. In the unimodal condition, no interaction is possible. Thus,

$$E[RT_{unimodal}] = E[S_1] + E[S_2|I^c] \quad (10)$$

and crossmodal interaction, CI , is defined as:

$$CI \equiv E[RT_{unimodal}] - E[RT_{crossmodal}] = \Pr[I] \cdot \Delta \quad (11)$$

Equation (11) expresses an important property of the TWIN model, the factoring of expected multisensory integration, into the probability of

interaction $\Pr[I]$ and the amount and sign of interaction (Δ). According to the assumptions, the first factor depends on the temporal configuration of the stimuli (SOA), whereas the second factor depends on their spatial configuration. Note that this *separation of temporal and spatial 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 closeness, whereas the amount and sign of interaction (Δ) is modulated by spatial proximity, reaching from enhancement for neighboring stimuli to possible inhibition for distant stimuli.

Predictions

The TWIN model makes a number of empirical predictions. First, the amount of crossmodal interaction should depend on the SOA between the stimuli. Indeed, the effect tends to be most prominent when there is some characteristic temporal asynchrony between the stimuli (Frens et al., 1995; Colonius and Arndt, 2001). Within the model, this simply means that a stimulus with faster peripheral processing has to be delayed (in terms of SOA) in such a way that the arrival times of both stimuli have a higher probability of falling into the window of integration.

Second, the probability of interaction, $\Pr[I]$, should depend on unimodal features that affect the speed of processing in the first stage, such as stimulus intensity or eccentricity. For example, if a stimulus from one modality is very strong compared to the intensity of the other stimulus, the peripheral processing speed of the former stimulus will increase, and the chances that both peripheral processes terminate within the time window will be smaller (assuming simultaneous stimulus presentations). The resulting low value of $\Pr[I]$ is in line with the empirical observation that a very strong target signal will effectively rule out any further reduction of RT by adding a stimulus from another modality (e.g., Corneil et al., 2002).

Finally, the amount of multisensory integration (Δ) and its direction (facilitation or inhibition) occurring in the second stage depend on

crossmodal features of the stimulus set, in particular 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.

Integration rule assumptions

More specific predictions require an explication of the rules governing the window-of-integration mechanism in specific task requirements.

- (5) *Focused attention task*: when the task is to orient toward the target stimulus ignoring stimuli from other modalities, the first stage terminates when the target peripheral process ends. Multisensory integration occurs only if the nontarget stimulus wins the race in the first stage opening a time window such that the termination of the target peripheral process is enclosed in the window.

In other words, in the focused attention situation, the window of integration is “opened” only by activity triggered by the nontarget stimulus, but first-stage duration is determined by the target stimulus processing time regardless of the identity of the winner of the race.

- (6) *Redundant target task*: when the task is to orient toward the first stimulus detected regardless of modality, the first-stage duration is defined by the winner’s peripheral processing time, and the window of integration is opened by whichever stimulus wins the race.

From these assumptions, further predictions concerning the effects of varying stimulus intensity follow. Take, for example, a focused attention task with a visual target and an auditory nontarget stimulus. Increasing the intensity of the visual stimulus will speed up visual peripheral processing (up to some minimum level) thereby increasing the chances 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, in contrast, 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.

Deriving the probability of interaction in TWIN

The race in the first stage of the model is made explicit by assigning independent non-negative random variables V and A to the peripheral processing times for the visual target and, for example, an auditory nontarget stimulus, respectively. With τ as SOA value and ω_I as integration window width parameter, the time window of integration assumption in the case of a focused attention experiment is equivalent to the (stochastic) event I , for example,

$$I = \{A + \tau < V < A + \tau + \omega\}$$

Thus, the probability of integration to occur, $P(I)$, 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.

Exponential distribution are postulated, for simplicity, for the peripheral processing time V for a visual target and A for an auditory nontarget with parameters λ_V and λ_A , respectively. That is,

$$f_V(t) = \lambda_V e^{-\lambda_V t}$$

$$f_A(t) = \lambda_A e^{-\lambda_A t}$$

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 $F_A(t)$.

In order to compute

$$\begin{aligned} P(I) &= Pr(A + \tau < V < A + \tau + \omega) \\ &= \int_0^\infty f_A(a) \{F_V(a + \tau + \omega) - F_V(a + \tau)\} da \end{aligned}$$

it is necessary to distinguish three cases for the magnitude of $\tau + \omega$ resulting in the following

expressions:

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

$$P(I) = \frac{\lambda_V}{\lambda_V + \lambda_A} e^{\lambda_A \tau} (-1 + e^{\lambda_A \omega})$$

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

$$P(I) = \frac{1}{\lambda_V + \lambda_A} \left\{ \lambda_A (1 - e^{-\lambda_V(\omega + \tau)}) + \lambda_V (1 - e^{\lambda_A \tau}) \right\}$$

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

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

Figures 2 and 3 depict the probability of integration as a function of the parameters λ and ω , respectively.

In the case of a redundant target experiment with both a visual and an auditory target, for example, the probability of integration is defined by both target processes finishing within the time window, that is,

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

and expressions similar to the above can be derived under the exponential assumption. More elaborate assumptions can be introduced with more than two target modalities.

Extending the TWIN model

The basic TWIN model can be extended in various ways. We consider two directions. First, in the focused attention case, empirical observations suggest that the nontarget may play an additional role as a warning cue modulating the RT under certain conditions. Second, the mechanism operating at the second processing stage determining the amount multisensory integration may be specified.

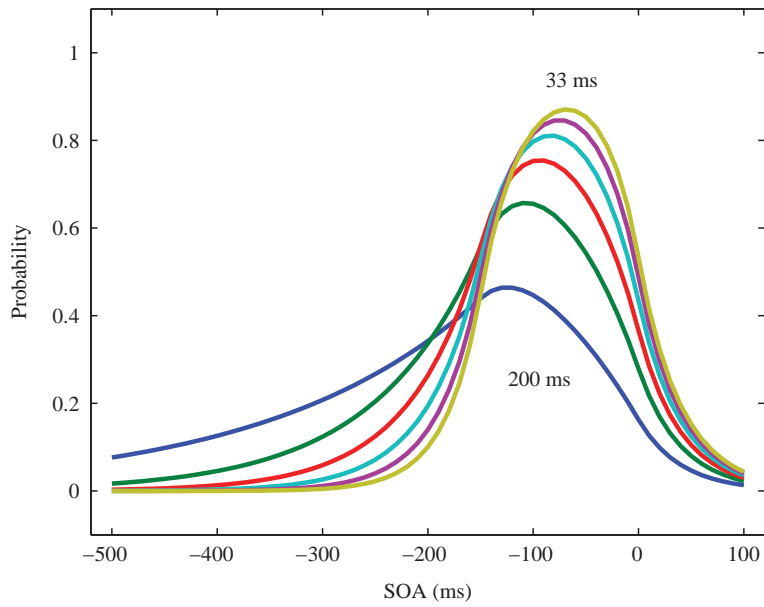


Fig. 2. Probability of integration as a function of the intensity parameter $\lambda_{nontarget}$ of the nontarget, ranging from $1/\lambda = 200$ ms to $1/\lambda = 33$ ms.

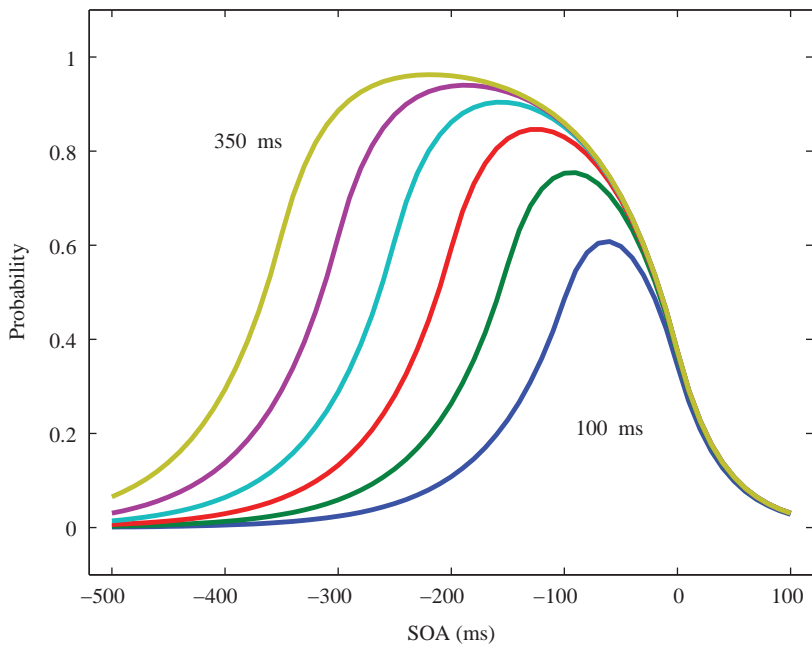


Fig. 3. Probability of integration as a function window width ω ranging from 100 ms to 350 ms in steps of 50 ms.

Warning mechanism: TWIN-W

Although estimates for the time window of integration vary somewhat across subjects and task specifics, the 200 ms width emerges in several studies (e.g., Eimer, 2001). In contrast, when the nontarget occurs at an earlier point in time (SOA of 200 ms or more before the target), a substantial decrease of RT compared to the unimodal condition has still been observed in our study (Diederich and Colonius, 2007a). This decrease, however, no longer depended on whether target and nontarget appeared at ipsi- or contralateral positions, thereby supporting the hypothesis that the nontarget plays the role of a *spatially unspecific* alerting or warning cue for the upcoming target whenever the SOA is large enough. Note that the hypothesis of increased crossmodal processing triggered by an alerting or warning cue had already been advanced in Nickerson (1973) who referred to it as “preparation enhancement.” In the eye movement literature, the effects of a warning signal have been studied primarily in the context of explaining the gap effect, that is, the latency to initiate a saccade to an eccentric target is reduced by extinguishing the fixation stimulus prior to the target onset (Reuter-Lorenz et al., 1991; Klein and Kingstone, 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 hand — 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.

- (7) *Assumption of warning mechanism:* if the nontarget wins the processing race in the first stage by a wide enough margin, then subsequent processing will be facilitated or inhibited without dependence on the spatial configuration of the stimuli.

The occurrence of warning depends on intramodal characteristics of the target and the nontarget

such as modality or intensity. For instance, an intense auditory nontarget may have a higher chance to win the race with a headstart (i.e., advantage) compared to a weak tactile nontarget.

The warning mechanism of the nontarget is triggered whenever the nontarget wins the race by a certain margin or headstart γ_A and, thus, its occurrence corresponds to the event:

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

The probability of warning to occur, $P(W)$, is a function of both τ and γ_A , and its value can be determined numerically as soon as the distribution functions of A and V have been specified. If the headstart γ_A is large enough for the integration window to close again, this implies:

$$\gamma_A > \omega \geq 0 \text{ and, therefore, } P(I \cap W) = 0$$

Assuming exponential distributions again, Fig. 4 shows the probability of integration for a window width of $\omega = 150$ ms and the probability of warning with various values of the headstart parameter γ .

The next step is to compute expected total RT for the unimodal and crossmodal conditions. From the two-stage assumption, total RT in the crossmodal condition can again be written as a sum of two random variables:

$$RT_{crossmodal} = S_1 + S_2 \quad (12)$$

where S_1 and S_2 refer, as before, to the first- and second-stage processing time, respectively. For the expected saccadic RT in the crossmodal condition then follows:

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

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

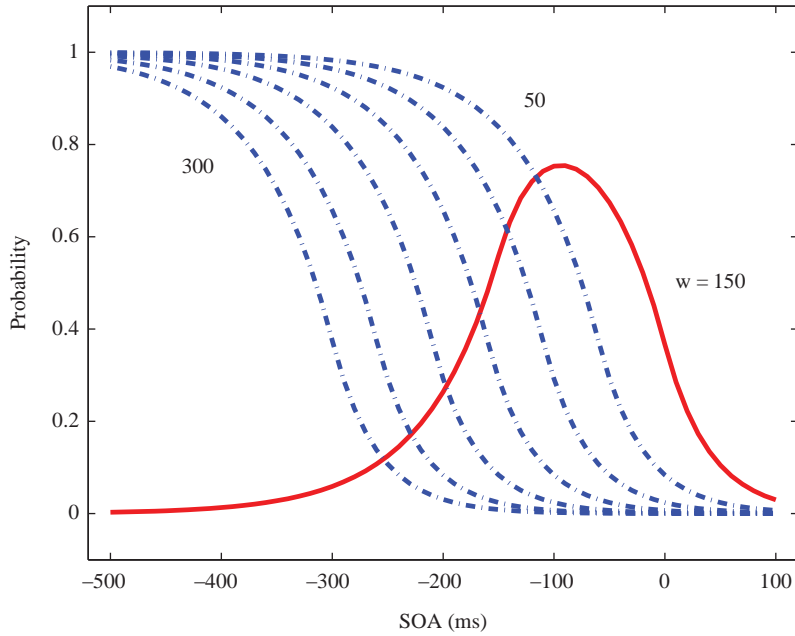


Fig. 4. Probability of integration (solid line) with window width $\omega = 150$ ms and probability of warning (dashed lines) with headstart γ ranging from 50 ms to 300 ms in steps of 50 ms.

conditioned on interaction occurring (I), warning occurring (W), or neither of them occurring ($I^c \cap W^c$), respectively (I^c , W^c represent the complement of events I , W). Setting:

$$\Delta \equiv E[S_2|I^c \cap W^c] - E[S_2|I]$$

$$\kappa \equiv E[S_2|I^c \cap W^c] - E[S_2|W]$$

this becomes:

$$E[RT_{crossmodal}] = E[S_1] + E[S_2|I^c \cap W^c] - P(I) \cdot \Delta - P(W) \cdot \kappa \quad (13)$$

In the unimodal condition, no integration or warning is possible. Thus,

$$E[RT_{unimodal}] = E[S_1] + E[S_2|I^c \cap W^c]$$

and we arrive at a simple expression for the combined effect of multisensory integration and

warning, crossmodal interaction (CI),

$$\begin{aligned} CI &\equiv E[RT_{unimodal}] - E[RT_{crossmodal}] \\ &= P(I) \cdot \Delta + P(W) \cdot \kappa \end{aligned} \quad (14)$$

Note 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.

The basic assumptions of TWIN-W imply that for a given spatial configuration and nontarget modality there are no sign reversals or changes in magnitude of Δ or κ across all SOA values. In contrast, both the probability of integration $P(I)$ and the probability of warning $P(W)$ do change with SOA. In particular, when the nontarget is presented very late relative to the target (large positive SOA), its chances 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, $P(I)$, is small. Therefore, the largest integration effects are expected for some mid-range SOA values. In contrast, 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 . Assuming exponential distributions once again, Figure 5 shows predicted mean saccadic reaction time (SRT) as a function of SOA under various conditions of warning and/or integration occurring.

It is interesting to note that this 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, for example, one would expect Δ to take on two different values, Δ_i and Δ_c , whereas $P(W) \cdot \kappa$, the expected nonspatial warning effect, should remain the same under both conditions. Subtracting the corresponding crossmodal interaction terms then gives, after canceling the warning effect terms (Eq. 14),

$$CI_i - CI_c = P(I) \cdot (\Delta_i - \Delta_c) \quad (15)$$

an expression that should yield the same qualitative behavior, as a function of SOA, as $P(I)$.

For further details and empirical tests of TWIN-W, we refer to Diederich and Colonius (2008a, b) and to Diederich et al. (2008).

Explicating the second stage of TWIN

Spatial stimulus configurations affect the second processing stage (see above). Spatially aligned stimuli may facilitate a response, whereas spatially disparate stimuli have no effect or may even inhibit a response. So far, no specific distributional assumptions about the second stage have been made. Colonius and Arndt (2001) assumed a normal distribution leading to an ex-Gaussian overall distribution. An alternative approach is to incorporate a sequential sampling process as proposed in Eq. (5).

Future directions for TWIN

The bidirectional links between action and perception could be made more explicit in the TWIN model framework. In fact, the separation of spatial and temporal factors determining crossmodal effects and the distinction between intersensory and warning effects as afforded by TWIN, could help to discern the perceptual and attentional consequences of a specific action. For example, moving the body in space, or even a simple turn of the head, may transform the spatial configuration of a crossmodal stimulus complex from a contralateral to an ipsilateral condition (relative to the observer) leading to specific changes in the amount of crossmodal effects, possibly transforming enhancement into inhibition or vice versa. In a similar manner, moving closer toward the source of an acoustical nontarget may increase its potential to act as an unspecific warning signal while, simultaneously, lowering its chances to play a role as a spatially specific part of a crossmodal stimulus configuration.

Another direction would be to extend the TWIN framework to account for choice response times and choice frequencies for two alternative choice problems as well (see Chapter 12: Embodied cognition of movement decisions: a computational modeling approach). To this end, the second stage of the TWIN model may be framed as a decision process. The idea is as follows. For a binary choice, for example between options A and B, a sequential sampling model is assumed. The options in this context could be actions such as reaching to the left or right, or moving the eyes to a target or not. Upon presentation of the choice options and completion of the first stage, the decision maker sequentially samples information from the stimulus display over time. The small increments of evidence sampled at any moment in time are such that they either favor option A or option B. Assume that a positive increment indicates an increment favoring option A and a negative increment is an increment favoring option B. The evidence is accumulated from one moment in time to the next by summing the current state with the new increment. This process continues until the magnitude of the cumulative

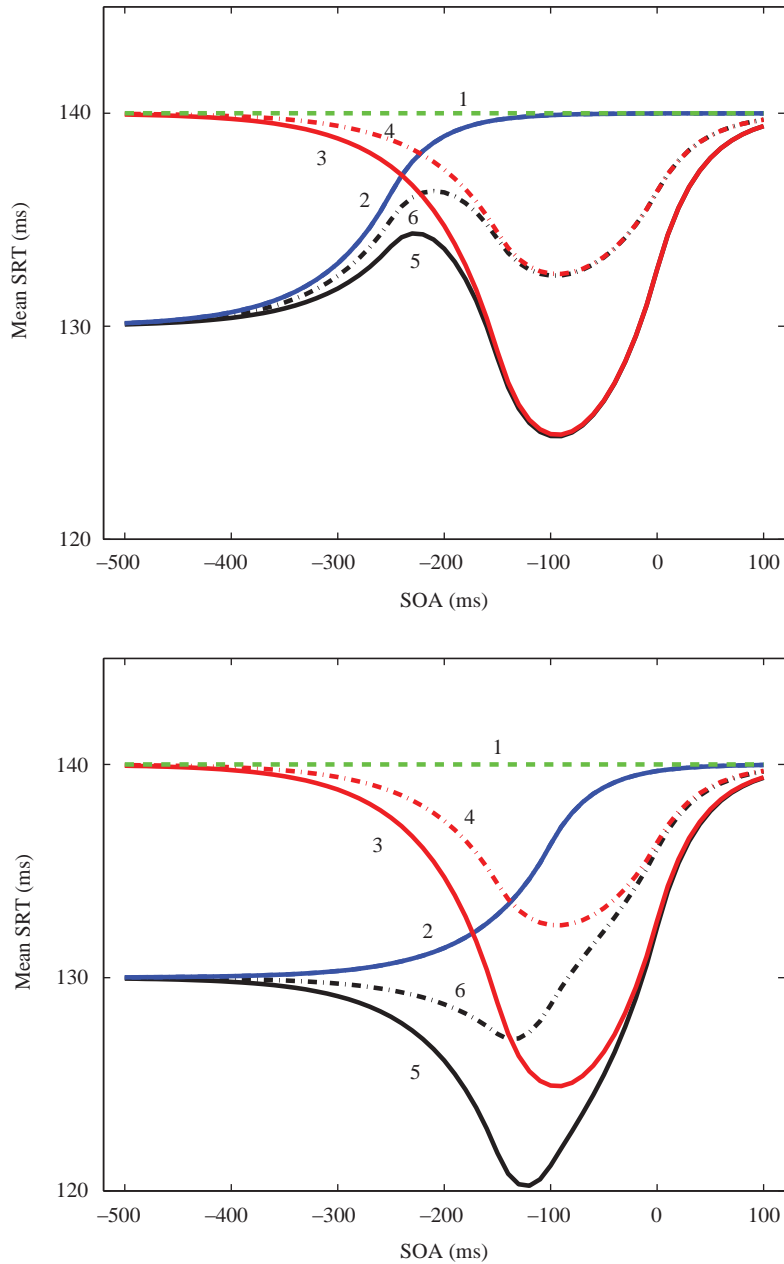


Fig. 5. Predicted mean saccadic reaction time (SRT) as a function of stimulus onset asynchrony (SOA). In the upper panel, $\gamma > \omega$; in the lower panel, $\gamma < \omega$. The horizontal (1) line indicates the predicted mean SRT to the target only. The remaining lines indicate the predicted mean SRT when target and nontarget are presented, as a function of SOA. Line 2 refers to the mean SRT when only warning takes place. Lines 3 and 4 refer to the predicted mean SRT when bimodal stimuli were presented ipsi- (3, solid) and contralaterally (4, dashed) and only integration occurs. The black lines refer to the predicted mean SRT when bimodal stimuli were presented ipsi- (5, solid) and contralaterally (6, dashed) and both integration and warning occur.

evidence exceeds a threshold criterion. The process stops and option A is chosen as soon as the accumulated evidence reaches a criterion value for choosing A, or it stops and chooses option B as soon as the accumulated evidence reaches a criterion value for choosing B. The probability of choosing A over B is determined by the accumulation process reaching the threshold for A before reaching the threshold for B. The criterion is assumed to be set by the decision maker prior to the decision task (e.g., see Diederich and Busemeyer, 2006; Diederich, 2008 for various models within this framework). Here, two criteria (two absorbing boundaries) of the process are assumed, one for initiating an A response and one for initiating a B response, while in the *multichannel diffusion* model stated above only one criterion (one absorbing boundary and one reflecting boundary) was assumed.

Within the context of the focused attention or redundant target paradigm, where subjects do not have to make a choice between alternative responses, an obvious interpretation for the options would be a choice between the decision to respond and the decision not to respond, given stimuli are presented at threshold levels or within a noisy background.

References

- Bell, A. H., Meredith, A., Van Opstal, A. J., & Munoz, D. P. (2005). Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology*, *93*, 3659–3673.
- Billingsley, P. (1979). *Probability and measure*. New York: Wiley.
- Calvert, G., Spence, C., & Stein, B. E. (Eds.). (2004). *Handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Colonus, H., & Arndt, P. (2001). A two-stage model for visual–auditory interaction in saccadic latencies. *Perception & Psychophysics*, *63*, 126–147.
- Colonus, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*, 1000–1009.
- Colonus, H., & Diederich, A. (2006). Race model inequality: interpreting a geometric measure of the amount of violation. *Psychological Review*, *113*, 148–154.
- Corneil, B. D., Van Wanrooij, M., Munoz, D. P., & Van Opstal, A. J. (2002). Auditory–visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology*, *88*, 438–454.
- Diederich, A. (1992). Probability inequalities for testing separate activation models of divided attention. *Perception & Psychophysics*, *52*, 714–716.
- Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*, 197–215.
- Diederich, A. (2008). A further test on sequential sampling models accounting for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, *70*, 229–256.
- Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: threshold-bound, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, *68*, 194–207.
- Diederich, A., & Colonius, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception & Psychophysics*, *50*, 83–86.
- Diederich, A., & Colonius, H. (2004a). Modeling the time course of multisensory interaction in manual and saccadic responses. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Diederich, A., & Colonius, H. (2004b). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404.
- 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. *Experimental Brain Research*, *179*, 43–54.
- Diederich, A., & Colonius, H. (2007b). Modeling spatial effects in visual–tactile saccadic reaction time. *Perception & Psychophysics*, *69*, 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. *Experimental Brain Research*, *186*, 1–22.
- Diederich, A., & Colonius, H. (2008b). When a high-intensity “distractor” is better than a low-intensity one: Modeling the effect of an auditory or tactile nontarget stimulus on visual saccadic reaction time. *Brain Research*, *1242*, 219–230.
- Diederich, A., Colonius, H., Bockhorst, D., & Tabeling, S. (2003). Visual–tactile spatial interaction in saccade generation. *Experimental Brain Research*, *148*, 328–337.
- Diederich, A., Colonius, H., & Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia*, *46*, 2556–2562.
- Eimer, M. (2001). Crossmodal links in spatial attention between vision, audition, and touch: evidence from event-related brain potentials. *Neuropsychologia*, *39*, 1292–1303.
- Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*, 802–816.

- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, *63*, 289–293.
- Howard, I. P., & Templeton, W. B. (1996). *Human spatial orientation*. London: Wiley.
- Klein, R., & Kingstone, A. (1993). Why do visual offsets reduce saccadic latencies? *Behavioral and Brain Sciences*, *16*, 583–584.
- Luce, R. D. (1986). *Response times: their role in inferring elementary mental organization*. New York: Oxford University Press.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and reading voices. *Nature*, *264*, 746–748.
- Meredith, M. A. (2002). On the neural basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, *14*, 31–40.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, *10*, 3215–3229.
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*, 640–662.
- Miller, J. O. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.
- Nickerson, R. S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement. *Psychological Review*, *80*, 489–509.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.
- Reuter-Lorenz, P. A., Hughes, H. C., & Fendrich, R. (1991). The reduction of saccadic latency by prior offset of the fixation point: an analysis of the gap effect. *Perception & Psychophysics*, *49*, 167–175.
- Ricciardi, L. M. (1977). *Diffusion processes and related topics in biology*. Berlin: Springer-Verlag.
- Ross, S. M., & Ross, L. E. (1981). Saccade latency and warning signals: effects of auditory and visual stimulus onset and offset. *Perception & Psychophysics*, *29*, 429–437.
- Schwarz, W. (1989). A new model to explain the redundant-signal effect. *Perception & Psychophysics*, *46*, 498–500.
- Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. *Journal of Mathematical Psychology*, *38*, 504–520.
- Sparks, D. L. (1986). Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus. *Physiological Reviews*, *66*, 116–177.
- Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*, R519–R521.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge MA: MIT Press.
- Todd, J. W. (1912). Reaction to multiple stimuli. In R. S. Woodworth (Ed.), *Archives of psychology*, No. 25. *Columbia contributions to philosophy and psychology* (Vol. XXI, No. 8). New York: The Science Press.
- Tuckwell, H. C. (1989). *Elementary applications of probability theory*. London: Chapman & Hall.
- Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia*, *45*, 598–607.
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of perception and human performance*, vol. 1: *sensory processes and perception* (pp. 25-1–25-36). New York: Wiley.
- Wood, T. W., & Recanzone, G. H. (2004). Cross-modal interactions evidenced by the ventriloquism effect in humans and monkeys. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes*. Cambridge, MA: MIT Press.