

Modeling the time course of multimodal interaction in manual and saccadic responses

Adele Diederich
International University Bremen

Hans Colonius
Universität Oldenburg

Running head: Modeling multimodal interaction

Contact author:

Hans Colonius
Phone: +49 441 798 5158
Fax: +49 441 798 5170
hans.colonius@uni-oldenburg.de

Introduction

Numerous effects of multisensory stimulation on perception and action have been described in the psychological literature over more than a century (cf. Welch and Warren, 1986). While many phenomena, like synesthetic experiences, can only be described at a qualitative level, other crossmodal effects are amenable to standard psychophysical measurement techniques. A prominent example is the "ventriloquism effect" where the presentation of a visual stimulus can shift the apparent location of a sound in the direction of the visual stimulus (see, e.g., Vroomen and de Gelder, this volume). However, as long as one must rely on the participants giving their subjective estimates, e.g., about sound localization, it is difficult to determine whether the observed crossmodal effect is a genuine perceptual effect or due to some kind of response bias. Similar qualifications apply to numerous studies trying to determine, e.g., the influence of a non-specific accessory stimulus from one modality on detection thresholds of a second modality. Such effects are generally small and unreliable, but the role of bias in threshold measurements can be assessed by using techniques from the theory of signal detection (e.g., Lovelace, Stein, and Wallace, 2002).

Response Speed as a Measure of Crossmodal Interaction

Another measure of perceptual sensitivity is the speed with which an observer is able to respond to the presence of a stimulus above threshold. Re-

response time (RT) 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). Although RT is measured on a physical (ratio) scale and seems not to be contaminated by judgment or response bias effects, the interpretation of RT data in the context of a specific experimental paradigm is subtle and requires a high level of technical skill that, fortunately, has been developed within experimental psychology (e.g., Van Zandt, 2002).

Response time analysis becomes most powerful in uncovering underlying processes when it is based on explicit quantitative hypotheses (Luce, 1986). Thus, the emphasis in this chapter is on mathematical models of response time 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.

One simple reason for the preponderance of response time as a measure of multimodal 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 response time 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 alone and the auditory alone (Hershenson, 1962). Interestingly, temporal integration rules similar to this one have been discerned for the responses at the level of individual multisensory neurons. 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, Nemitz, and Stein, 1987). This overlap is often not correlated with simultaneous stimulus presentation nor with differences in stimulus presentation that match their latency differences. Since multisensory SC neurons project to premotor areas of the brain stem 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 (cf. Colonius and Diederich, 2002a).

Crossmodal Paradigms

Experimental paradigms to elicit a response to a crossmodal stimulus set differ with respect to the instructions given to the participant, and these differences must be taken into account by models of response speed. In the *redundant target paradigm* stimuli from different modalities are presented simultaneously or with certain interstimulus intervals, and the subject is instructed to respond to whatever stimulus is detected first. The participant is not asked to identify the modality that elicited his or her response and often is not able to. 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 specific target modality, the visual, say. If in some of the trials no stimulus from the target modality is presented, these trials are called *catch* trials.

In all paradigms participants are prone to make certain types of errors: *anticipation errors* are responses made before any stimulus appears, and *misses* occur if no response is made at all. These two types of errors can often be attributed to slips of attention, are minimized by training, and are

typically not part of the modeling effort. Another type of error are responses made in catch trials. Their frequency depends on whether the participant is explicitly instructed to respond as fast as possible and not to worry about avoiding erroneous catch trial responses or, alternatively, to respond at a speed level that minimizes these errors. In the models to be discussed here, no mechanism to explain catch trial responses is adopted since catch trial error frequency is typically kept at a negligible level (less than 1 or 2%).

Manual vs. Saccadic Responses

The ubiquitous mode of response speed assessment has been to measure the time it takes to press a button, or to release it, by a finger or foot movement (cf. Donders 1868/1969). With the advance of modern eye movement registration techniques the measurement of *saccadic reaction time*, i.e., the time from the presentation of the target stimulus to the beginning of the eye movement, is gaining in popularity. As the superior colliculus is an important site of oculomotor control (e.g., Munoz and Wurtz, 1995), measuring saccadic RT is an obvious choice to study behavioral consequences of multisensory integration.

An important distinction in measuring response speed is whether a correct response requires the participant to make a choice. For example, a participant may be asked to respond to a visual stimulus with the left hand and to an auditory stimulus with the right hand, or to respond to a stimulus on the right, irrespective of modality, with the right hand and to a stimulus on the

left with the left hand (or vice versa). In any event, the possible number of stimulus-response mappings is somewhat restricted (even if responding with different fingers is employed, a procedure that adds possibly unwanted compatibility effects) as long as manual responses are employed.

For saccadic eye movements, however, this restriction does not apply given that the possible number of target positions in space is limited only by the spatial resolution of the oculomotor system. Thus, saccadic responses are ideally suited for studying the spatial rules of crossmodal interaction, in parallel to the determination of the temporal rules. Eye movement registration is possible under both the redundant target and the focused attention paradigm, and participants can be asked to move their eyes to either visual, auditory, or somatosensory targets. Nevertheless, the model builder should be aware of certain idiosyncrasies due to the simple fact that the ocular system is dominated by the visual system. For example, it is well known that saccades to visual targets have a higher level of accuracy than those to auditory or somatosensory stimuli (see Van Opstal, and Munoz, this volume).

Preview

Current models of response time can roughly be divided into two classes, *separate activation models* and *coactivation models* (a terminology suggested in Miller, 1982). Some models of either type will be described in the subsequent sections including a discussion of their empirical merits, but for a

complete coverage of experimental results we must refer to the literature. Finally, we present the recently developed *time-window-of-integration model* that combines features of both model types (cf. Colonius and Diederich, 2002b).

Separate Activation Models

Separate activation models, also known as *race models*, assume (1) that presenting a multimodal 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 reaction time to multimodal 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

$$E[\min(T, V, A)] \leq \min(E[T], E[V], E[A]), \quad (1)$$

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

To illustrate, consider a visual and an auditory stimulus, say, with processing time distributions f_V , f_A , and f_{VA} for the visual, auditory, and the bimodal (visual–auditory) condition, respectively. Suppose, the subject consistently responds faster to the unimodal auditory stimulus than to the unimodal visual. In that case, distribution f_A would lie completely left of distribution f_V (no overlap). Thus, in the bimodal condition the subject’s reaction time would be determined by the time to process the auditory stimulus only. More realistically, however, f_V and f_A do overlap and, thus, some of the long reactions to the auditory stimulus will be replaced by shorter reactions to the visual stimulus in the bimodal condition. Therefore, mean bimodal RT will tend to be smaller than mean RT to the auditory stimulus. The more the distributions overlap the more often long reactions to the auditory stimulus will be replaced by shorter reactions to the visual stimulus resulting in smaller bimodal mean RT. Figure 1 illustrates this effect.

[Figure 1 about here]

Maximal overlap and thus maximal statistical facilitation is expected if presentation of the faster processed stimulus (e.g., the auditory stimulus) is delayed by an amount of time equal to the difference in mean RT in the unimodal stimulus conditions, i.e., when *physiological synchrony*, rather than stimulus presentation synchrony, is achieved.

Testing Separate Activation Models

Assuming independent Gaussian distributions for the processing times Raab's model (Raab, 1962) fell slightly short of predicting the facilitation observed in Hershenson (1962). Gielen, Schmidt, and van den Heuvel (1983) generalized Raab's model by employing the observed unimodal distributions to estimate the minimum distribution in the bimodal conditions. Nevertheless, statistical facilitation alone could not account for the facilitation observed in their data.

It should be noted that separate activation models need not assume processing times in the multimodal conditions to be statistically independent. Specifically, the amount of statistical facilitation increases with negatively dependent processing times because longer processing times of one modality then tend to occur more often with shorter processing times of the other modality (cf. Colonius, 1990).

The prevailing test of separate activation models was proposed by Miller (1982). It is based on a simple inequality (Boole's inequality, cf. Billingsley, 1979) holding for arbitrary events E_1 and E_2 in a probability space,

$$P(E_1 \cup E_2) \leq P(E_1) + P(E_2). \quad (2)$$

Identifying E_1 with the event $\{V \leq t\}$, E_2 with $\{A \leq t\}$ yields $E_1 \cup E_2 = \{\min(V, A) \leq t\}$, so that Eq. 2 becomes, for all t ,

$$P(\min(V, A) \leq t) \leq P(V \leq t) + P(A \leq t). \quad (3)$$

Assuming that the observable reaction times in the unimodal and bimodal conditions, RT_V , RT_A , and RT_{VA} , follow the same distribution as V , A , and $\min(V, A)$, resp., puts an upper bound on the facilitation produced by bimodal stimuli (cf. Miller, 1982),

$$P(RT_{VA} \leq t) \leq P(RT_V \leq t) + P(RT_A \leq t) \quad \text{for all } t. \quad (4)$$

If this inequality is violated for some value of time t , all separate activation models, whether statistical independence among the channel processing times is assumed or not, can be rejected.

Miller's test has become the standard tool to assess whether statistical facilitation suffices to explain the response speed-up of bimodal stimulus presentations (Townsend and Nozawa, 1995). Nevertheless, there are a number of caveats to be aware of when employing this test. First, the inequality is not yet amenable to sound statistical testing. The problem is that its right hand side is not a probability distribution (it converges toward 2 for large t values). For the same reason, violations can only be expected for small enough values of t . Second, the inequality is based on the assumption of *context independence* (Townsend and Ashby, 1983; Colonius, 1990). While different from statistical independence, this means that the distribution of V (or A) should be the same whether the stimulus is presented in the uni- or in the bimodal condition, and it is not clear how to test this empirically. Third, it should be noted that non-violation of the inequality does not provide evidence in favor of separate activation models (Colonius, 1986, 1990;

Diederich and Colonius, 1987; Miller, 1982; Ulrich and Giray, 1986).

The amount of violation of the inequality can also be utilized as an indicator of the amount of statistical facilitation that is present in the data. Subtracting the right hand side of Eq. 4 from the left hand side and plotting the difference as a function of t yields a curve that takes on positive values wherever the inequality is violated. Thus, the area under the (positive part of) the difference function is a measure of the degree of violation.

As an illustrating example, Fig. 2 presents saccadic RT data from a visual-auditory focused attention study where both spatial disparity between visual target and auditory accessory and intensity of the auditory stimulus were orthogonally combined at 3×3 levels (Arndt and Colonius, 2003).

[Figure 2 about here]

The area indicating violation is about constant within each row, with obvious violations for coincident stimuli (0°), whereas for disparate stimuli very little (25°) or no violation at all (50°) is exhibited. It is noteworthy that the intensity level of the auditory accessory has no discernible influence on the amount of violation (for an explanation see below).

With the same rationale as for before, Ineq. 4 can be extended to test separate activation models with trimodal (e.g., tactile-visual-auditory) stimulus sets:

$$P(RT_{TVA} \leq t) \leq P(RT_T \leq t) + P(RT_V \leq t) + P(RT_A \leq t), \quad (5)$$

where RT_{TVA} refers to the observable reaction time to the tactile-visual-

auditory stimulus. This inequality puts an upper bound on the facilitation produced by trimodal stimuli. However, since it is the sum of three probabilities approaching 3 for increasing values of t , this bound is not likely to be of much use in testing separate activation. Interestingly, the upper bound can be sharpened in various ways. Let E_1, \dots, E_m be m arbitrary events in some probability space. Worsley (1982) proved that

$$P\left(\bigcup_{i=1}^m E_i\right) \leq \sum_{i=1}^m P(E_i) - \sum_{i=1}^{m-1} P(E_i \cap E_{i+1}) \quad (6)$$

holds. Taking $m = 3$ and identifying E_1 , E_2 , and E_3 with the events $\{T \leq t\}$, $\{V \leq t\}$, and $\{A \leq t\}$ for processing a tactile, visual, and auditory stimulus before time t , resp., yields $P(E_1 \cup E_2 \cup E_3) = P(\min(T, V, A) \leq t)$. As shown in Diederich (1992), this leads to three different inequalities relating the observable reaction time distributions in uni-, bi- and trimodal conditions. One of them is

$$\begin{aligned} P(RT_{TVA} \leq t) &= P(\min(T, V, A) \leq t) \\ &\leq P(\min(T, V) \leq t) + P(\min(V, A) \leq t) - P(RT_V \leq t) \\ &= P(RT_{TV} \leq t) + P(RT_{VA} \leq t) - P(RT_V \leq t), \end{aligned}$$

and the other two follow by combining other pairs of modalities analogously. Note that the upper bound should be sharper now since it approaches 1 for large enough t values. Moreover, taking at each t the minimum of all three upper bounds may lead to an even sharper upper bound.

This extension of the inequality test to three modalities is illustrated here with data from an experiment where a tactile stimulus was presented first,

followed by a visual stimulus after τ_1 ms, followed by an auditory stimulus $\tau_2 - \tau_1$ ms later (Diederich, 1995). Writing $RT_{T\tau_1 V(\tau_2 - \tau_1)A}$ for the reaction time of this trimodal stimulus, $RT_{T\tau_1 V}$ for the reaction time to a tactile stimulus followed by a visual stimulus τ_1 ms later, and so on, the following inequality test obtains:

$$P(RT_{T\tau_1 V(\tau_2 - \tau_1)A} \leq t) \leq P(RT_{T\tau_1 V} \leq t) + P(RT_{V(\tau_2 - \tau_1)A} \leq t) - P(RT_V + \tau_1 \leq t). \quad (7)$$

Figure 3 depicts the trimodal RT data (stars) together with (1) the upper bound of Ineq. 7 obtained from the sum of two bimodal RT data minus the unimodal RT data (circles) and (2) the upper bound of Ineq. 5 obtained from the sum of three unimodal RT data (diamonds). Separate activation would predict an ordering of the curves reverse to what is seen (up to about 160 ms), so these data present further evidence against this class of models.

[Figure 3 about here]

Coactivation Models

Given the frequent empirical failure of separate activation models, an alternative model type has been considered to explain intersensory facilitation. *Coactivation models* assume that activation raised in different sensory channels by presenting multimodal stimuli is combined to satisfy a single criterion

for response initiation. Coactivation models predict faster average reaction time to multiple stimuli compared to single stimuli because the combined activation reaches that criterion faster.

An early version of this concept is the *energy summation* hypothesis according to which the energy from the different sensory modalities, determined by stimulus intensities, is assumed to be added during some peripheral stage of processing. Bernstein and collaborators (Bernstein et al. 1970) presented an "energy integration model" according to which the intensities of visual and auditory stimuli integrate at some common point in the nervous system. Presenting two different stimuli amounts to the same as increasing the intensity of either stimulus. Empirical evidence with respect to the energy summation hypothesis turned out mixed. For example, decreasing auditory intensity often only slightly affected facilitation. Moreover, the fact the spatial configuration of the stimuli has a clear effect on facilitation is not easily reconcilable with the hypothesis (for a review, see Nickerson, 1973). As an alternative to energy summation Nickerson proposed the *preparation-enhancement model*. Here one stimulus plays an alerting role at many of the processing stages, so that the affected stages are terminated more quickly and the response comes earlier. The fact that these models are formulated mainly at a descriptive level makes testing them rigorously a difficult task. The two coactivation models considered next have a clear mathematical basis.

Superposition Models

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, say, 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, i.e., 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, $N_V(t)$, $N_A(t)$, $N_T(t)$, for a visual, auditory, and tactile stimulus, respectively. Presentation of a multimodal stimulus triggers all counters to start registering their sequence of modality-specific events, and the counter which first reaches a preset criterion initiates the response. Obviously, under this assumption we are back 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 cross-modal 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 4 illustrates this intuition.

[Figure 4 about here]

To compute the distribution of the (random) *waiting time* S_c for the c th count to occur note that (cf. Cinlar, 1975)

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

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 then simply is c/λ .

Superposition models represent reaction time 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, e.g., 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 stimulus, resp., the expected waiting time for the c th count to occur when all three stimuli are presented is

$$E[S_c | trimodal] = \frac{c}{\lambda_T + \lambda_V + \lambda_A} \quad (8)$$

At the level of mean reaction times, 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 50dB and a 70dB tone with $\lambda_{50} < \lambda_{70}$,

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

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

[Figure 5 about here]

Figure 5 presents the fit of the model to data from a redundant target experiment with trimodal and bimodal stimuli including various stimulus onset asynchrony (SOA) conditions (Diederich, 1992). In particular, the tactile and

the visual stimulus were presented with SOA τ_1 indicated at the abscissa. For the trimodal stimulus condition the auditory stimulus was presented τ_2 ms after the tactile stimulus (upper panel: $\tau_2 = 40$ ms, lower panel: $\tau_2 = 60$ ms). As predicted, mean RT to trimodal stimuli is shorter than to bimodal stimuli. Overall, at the level of the means the model gave a very satisfactory fit for this rather large set of data. In particular, it was 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), was 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.

On the other hand, the direct representation of stimulus intensity by the intensity parameter falls short of two ubiquitous empirical observations. First, the model predicts that increasing stimulus intensity should lead to ever faster responses, without being able to account for any saturation effects. Second, it is not clear how the observation of "inverse effectiveness", according to which crossmodal facilitation is strongest when stimulus strengths are weak, can be predicted by the superposition model. These shortcomings of superposition models have led to considering still another version of the coactivation idea.

The Multi-Channel Diffusion Model

This 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 multi-channel 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 infinitely small amount. Each presentation of a stimulus triggers the realization of a function describing the course of activation over time, called a *trajectory*.

[Figure 6 about here]

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:

$$\mu(x) = \delta - \gamma \cdot x. \quad (9)$$

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 quickly approach

the criterion level to initiate a response. Note, however, that for positive values of γ (the *decay parameter*) the drift $\mu(x)$ decreases the faster the larger the activation level x becomes, i.e., the closer activation gets to the criterion level. This is responsible for the trajectories to level off rather than to increase linearly over time (Figure 6, upper panels). Moreover, when the stimulus signal is switched off, the drift becomes negative and 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 multimodal stimuli the δ values corresponding to the unimodal stimuli are added:

$$\mu(x) = (\delta_T + \delta_V + \delta_A) - \gamma \cdot x \quad (10)$$

Figure 6 (lower panels) illustrates the effect of the drift change when modalities are added after certain SOAs. It is not difficult to see that the multi-channel 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).

Superposition vs. Diffusion Models

While the data base to evaluate the relative merits of both types of coactivation models is not large enough for a definite answer, the diffusion models seems to be more flexible to deal 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, inverse effectiveness could be introduced by an appropriate elaboration of the drift function. For example, replacing the term $\delta_T + \delta_V + \delta_A$ in Eq. 10 by

$$(\delta_T + \delta_V + \delta_A)[1 + (\delta_T^{max} - \delta_T)(\delta_V^{max} - \delta_V)(\delta_A^{max} - \delta_A)], \quad (11)$$

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

The Time-Window-of-Integration (TWIN) Modeling Scheme

One feature of all models considered so far, both of separate activation or coactivation type, is their complete functional symmetry 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 experimental paradigms like the focused attention paradigm. The modeling scheme proposed in the following incorporates explicit assumptions about possible target/non-target modality distinctions.

While there is increasing evidence for a complex network of largely parallel neural subprocesses underlying performance even in simple crossmodal tasks (Driver and Spence, 2000), the initial separation of the afferent pathways for the different sensory modalities suggests that one can distinguish at least two serial stages of the entire saccadic reaction time: 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 additional weak assumptions some interesting and empirically testable predictions can be derived from this simple setup.

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 multimodal stimulus complex.

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

Second Stage Assumption: The *second stage* comprises neural integration of the input and preparation of an ocular motor response. Crossmodal interaction manifests itself in an increase or decrease of second stage processing time.

Distinguishing only between two stages is clearly an oversimplification. But note 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.

Time-Window-of-Integration Assumption: Crossmodal interaction occurs only if the peripheral processes of the first stage all terminate within a given time interval, the "window of integration".

This window of integration acts like a filter determining whether the 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 crossmodal interaction to occur. The reason is that crossmodal interaction 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 crossmodal interaction to occur even for rather distant stimuli (of different modality) as long as they fall within the time window. Such interaction

will typically be an inhibition or only a small facilitation. Note that this arrangement affords the organism 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 Crossmodal Interaction in the TWIN Model

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

$$RT_{multimodal} = W_1 + W_2, \quad (12)$$

where W_1 and W_2 refer to the first and second stage processing time, respectively. Let I denote the event that crossmodal interaction occurs, having probability $\Pr(I)$. For the expected saccadic reaction time in the multimodal condition then follows:

$$\begin{aligned} \mathbb{E}[RT_{multimodal}] &= \mathbb{E}[W_1] + \mathbb{E}[W_2] \\ &= \mathbb{E}[W_1] + \Pr[I]\mathbb{E}[W_2|I] + (1 - \Pr[I])\mathbb{E}[W_2|\text{not-}I] \\ &= \mathbb{E}[W_1] + \mathbb{E}[W_2|\text{not-}I] - \Pr[I](\mathbb{E}[W_2|\text{not-}I] - \mathbb{E}[W_2|I]), \end{aligned}$$

where $\mathbb{E}[W_2|I]$ and $\mathbb{E}[W_2|\text{not-}I]$ denote the expected second stage processing time conditioned on interaction occurring (I) or not occurring ($\text{not-}I$), respectively. Putting $\Delta \equiv \mathbb{E}[W_2|\text{not-}I] - \mathbb{E}[W_2|I]$, this becomes

$$\mathbb{E}[RT_{multimodal}] = \mathbb{E}[W_1] + \mathbb{E}[W_2|\text{not-}I] - \Pr[I] * \Delta. \quad (13)$$

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

$$E[RT_{unimodal}] = E[W_1] + E[W_2|\text{not-}I], \quad (14)$$

and

$$E[RT_{unimodal}] - E[RT_{multimodal}] = \Pr[I] * \Delta \quad (15)$$

Predictions

The TWIN model makes a number of empirical predictions. First, the amount of crossmodal interaction should depend on the stimulus onset asynchrony (SOA) between the stimuli. Indeed, the effect of crossmodal interaction tends to be most prominent when there is some characteristic temporal asynchrony between the stimuli (Frens et al., 1995). Within the model this simply means that a stimulus with faster peripheral processing has to be delayed 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, like stimulus intensity or eccentricity. For example, if a stimulus from one modality is very strong compared to the other stimulus' intensity, the chances that both

peripheral processes terminate within the time window are small (assuming simultaneous stimulus presentations). The resulting low value of $\text{Pr}(I)$ is in line with the empirical observation that a very strong target signal will effectively suppress any interaction with other modalities.

On the other hand, the principle of "inverse effectiveness", according to which crossmodal facilitation is strongest when stimulus strengths are weak or close to threshold level (Meredith and Stein 1986), can be accommodated in the model by adjusting the width of the time window: for low-level stimuli the window should become larger so as to increase the likelihood of cross-modal integration.

Finally, the amount of crossmodal interaction (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. More specific predictions require an explication of the rules governing the window-of-integration mechanism in specific task requirements to be discussed next.

Integration Rule Assumptions

Focused Attention Task: When the task is to orient toward the target stimulus ignoring stimuli from other modalities, the first

stage duration is determined by the target peripheral process, but crossmodal integration is effective only if the non-target stimulus wins the race in the first stage.

In other words, in the focused attention situation, the window of integration is "opened" only by activity triggered by the non-target stimulus.

Redundant Target Task: When the task is to orient toward the first stimulus detected no matter of which modality, the first stage duration is determined 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 non-target stimulus. Increasing the intensity of the visual stimulus will increase the chances for the visual target to win the race. Thus, the probability that the window of integration opens decreases predicting less crossmodal interaction. This is again in line with the observation that a very strong target signal will tend to suppress interaction with any other modality. Increasing the intensity of the non-target auditory stimulus, however, 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 crossmodal interaction to occur on average.

Separating Spatial and Temporal Factors of Integration

Expected crossmodal interaction is defined as the difference between mean response time in the unimodal and the crossmodal condition. An important property of the TWIN model is the factoring of expected crossmodal interaction, i.e., expected crossmodal interaction is simply the product of the probability of interaction $\text{Pr}(I)$ and the amount and sign of interaction (Δ) (see Eq. 15). According to the assumptions, the first factor depends on the temporal configuration of the stimuli (stimulus onset asynchrony, 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 crossmodal interaction hinges upon the stimuli to occur close in time, whereas the amount and sign of interaction (Δ) is modulated by spatial proximity, reaching from enhancement for close stimuli to possible inhibition for distant stimuli.

TWIN Model for Visual-Tactile Interaction in Focused Attention

To illustrate the modeling scheme, a TWIN model application to a visual-tactile interaction experiment under focused attention is sketched here (cf. Colonius and Diederich, 2002b, for details). Subjects were asked to make a saccade as quickly and as accurately as possible toward a visual stimulus appearing randomly left or right of the fixation point. They were instructed to ignore a tactile accessory stimulus that, in bimodal trials, was applied at

different hand positions ipsi- or contralateral to the visual target. Stimulus onset asynchrony (SOA) between visual and tactile stimulus was -100 , -50 , 0 , and 50 ms.

With V denoting visual peripheral processing time, T tactile peripheral processing time, τ stimulus onset asynchrony, and ω the width of the integration window, crossmodal interaction depends on the event I that the tactile stimulus wins the race in the first stage and that visual peripheral processing terminates within the window of integration,

$$I = \{T + \tau < V < T + \tau + \omega\}.$$

For a quantitative prediction, the peripheral processes in the first stage are assumed to have stochastically independent exponentially distributed durations. The exponential assumption is motivated by mathematical simplicity and, together with a Gaussian distribution assumption for second stage processing time, results in an Ex-Gaussian distribution from that has been demonstrated to be a reasonably adequate description for many empirically observed reaction time data (cf. Van Zandt, 2002). The first stage duration is determined by the target peripheral process of random duration V , say, yielding $E[W_1] = E[V] = 1/\lambda_V$ (λ_V denotes the intensity parameter of the exponential distribution of V). From the assumptions stated in the last section. Straightforward calculation yields

$$P[I] = \frac{\lambda_T}{\lambda_T + \lambda_V} \{ \exp[-\lambda_V \tau] - \exp[-\lambda_V(\tau + \omega)] \}, \quad (16)$$

where λ_T refers to the tactile intensity parameter. It is obvious from Eq. (16)

that the probability of interaction increases both with λ_T and the window width ω , as it should. Expected saccadic reaction time then is (cf. Colonius and Diederich, 2002b)

$$E[RT_{multimodal}] = 1/\lambda_V + \mu - \frac{\Delta \lambda_A}{\lambda_A + \lambda_V} \{\exp[-\lambda_V \tau] - \exp[-\lambda_V(\tau + \omega)]\}$$

where $\mu = E[W_2|\text{not-}I]$, the mean duration of the second stage when no interaction occurs.

The choice of the second stage distribution is irrelevant as long as only mean latencies are considered. For predictions of the entire saccade latency distribution it should be noted, however, that due to conditioning on the event of interaction I the two stage durations W_1 and W_2 are not stochastically independent. For the model version considered in this section, it can be shown that they are negatively dependent if Δ is positive: in any given trial, whenever the visual peripheral process ($V \equiv W_1$) is relatively slow, the tactile peripheral process has a better chance of winning the race and opening the integration window, thus increasing the likelihood of facilitation in the second stage, and vice versa.

[Figure 7 about here]

The data from one of the subjects (Figure 7) show that saccadic reaction time to a visual target presented randomly left or right from fixation was reduced by up to 22% when accompanied by a spatially non-predictive tactile stimulus. This facilitation effect was larger for spatially aligned visual-tactile stimulus configurations than for contralateral presentation, and it increased

with stimulus eccentricity (20° vs. 70°). Moreover, responses were the faster the earlier the tactile stimulus was presented (in a range of the tactile preceding the visual by 100 ms to following it by 50 ms). These results extend previous findings to non-simultaneous visual-tactile stimulus presentations (Diederich, Colonius, Bockhorst, & Tabeling, 2003). Obviously, the model captures the main features of the data at the level of the means. Mean saccadic reaction time is increasing with SOA for all stimulus configurations, except for the 70° contralateral condition where monotonicity is reversed due to inhibition, consistent with the model. The estimate for the window width was 200 ms.

Conclusions

For an evaluation of the TWIN model it is important to realize that it is not meant to mirror the processes at the level of an individual neuron. There are many different types of multisensory convergence occurring in individual neurons (for a recent review, see Meredith, 2002) and some of their activities are consistent with the TWIN assumptions while others are not. For example, for certain neurons in cat SC, given a visual stimulus is always presented in the visual receptive field of the bimodal (visual-auditory) neuron, an auditory stimulus presented outside its receptive field will generate only response depression (or no interaction) regardless of the SOA between the stimuli (Stein & Meredith, 1993, p. 140). This accords nicely with the separation of temporal and spatial factors postulated by the TWIN model. However, in

other visual-auditory neurons changing the temporal order and interval between the stimuli can change enhancement to depression even if their spatial arrangement is left invariant (Meredith, Nemitz, & Stein, 1987). This latter type of behavior could only be accounted for in TWIN if two stimuli, after having passed the filter of the temporal window of integration, could still carry over information about their temporal distance to the subsequent convergence stage. Moreover, besides the common excitatory-excitatory type of multisensory convergence there is evidence for an excitatory-inhibitory type of neural circuit by which inputs from one modality inhibit those from the other (cf. Meredith, 2002). These circuits could play a specific role in focused attention situations (ibid, p. 37). In any event, at this stage of development the TWIN model operates at the level of behavioral data resulting from the combined activity of a possibly large number of neurons and from the specific task instructions, so that the existence of different types of multisensory convergence in individual neurons does not provide a strong modeling constraint.

Since stimulation from different modalities like vision and touch cannot interact (e.g., on the retina), the main assumption of the two-stage model, the existence of a first stage of parallel independent modality-specific activations in the afferent pathways seems non-controversial. It refers to a very early stage of processing where detection of the stimuli, but possibly no "higher" processes like localization and identification, take place. Note that the two-stage assumption does not preclude the possibility of interaction between

modality-specific pathways, nor between modality-specific and crossmodal areas, at a later stage. In fact, there is increasing evidence that crossmodal processing does not take place entirely in feedforward convergent pathways but that it can also modulate early cortical unisensory processing, as suggested by recent functional magnetic resonance imaging (fMRI) and ERP studies (see Macaluso, Frith, & Driver, 2000, and Laurienti, Burdette, Wallace, Yen, Field, & Stein, 2002, for fMRI; Molholm, Ritter, Murray, Javitt, Schroeder, & Foxe, 2002, for ERP).

Even if certain assumptions of the TWIN model eventually turn out to be inconsistent with experimental results, an advantage of this modeling framework is that it facilitates the statement of clear-cut hypotheses about the rules of crossmodal integration. A case in point the hypothesis of "restricted influence": it holds that unimodal stimulus properties, like stimulus intensity, do not have a direct effect on processing in the second stage. Of course, unimodal properties may influence the resulting crossmodal integration indirectly by modulating the opening of the integration window. Note that the restricted influence hypothesis is a strengthening of the (obvious) TWIN model assumption that unimodal properties affect the race among the modalities in the first stage of processing. This is consistent with the focused attention study mentioned before (Arndt and Colonius, 2003). In a similar vein, in a redundant target experiment Corneil, Wanrooij, Munoz, & Van Opstal (2002) did not find an effect of auditory signal-to-noise ratio on the amount of crossmodal interaction.

Further work on the TWIN model should include specification of the second stage mechanisms, in particular with respect to the spatial stimulus configuration effects. There is a large data base on receptive field properties of multisensory neurons available now (cf. Kadunce et al., 2001), and connecting these with behavioral data via an appropriate elaboration of the TWIN model should be a challenging task.

References

- Arndt, A., and H. Colonius, (2003) Two separate stages in crossmodal saccadic integration: Evidence from varying intensity of an auditory accessory stimulus. Exp. Brain Res., in press.
- Billingsley, P., 1979. Probability and measure, New York: John Wiley and Sons.
- Bernstein, I. H., R. Rose, and V. M. Ashe, 1970. Energy integration in intersensory facilitation. J. Exp. Psychol., 86:196-203.
- Colonius, H., 1986. Measuring channel dependence in separate activation models. Perc. Psychophysics, 40:251-255.
- Colonius, H. 1990. Possibly dependent probability summation of reaction time. J. Math. Psychol., 34:253-275
- Colonius, H., and P. Arndt, 2001. A two-stage model for visual-auditory interaction in saccadic latencies. Perc. Psychophysics, 63:126–147.
- Colonius, H., and A. Diederich, 2002a. A maximum-likelihood approach to modeling multisensory enhancement, in Advances in Neural Information Processing Systems 14, (T. G. Dietterich, S. Becker, and Z. Ghahramani, eds.), Cambridge: MIT Press, pp. 181-187.
- Colonius, H. and A. Diederich, 2002b. A stochastic model of multimodal integration in saccadic responses, in Dynamic Perception, (R. P. Würtz

and M. Lappe, eds.), Berlin: Akademische Verlagsgesellschaft, pp. 321–326.

Colonius, H., and Diederich, A., submitted. The time-window-of-integration model: Visual-tactile interaction in saccade generation.

Corneil, B. D., M. Van Wanrooij, D. P. Munoz, and A. J. Van Opstal, 2002. Auditory-visual interactions subserving goal-directed saccades in a complex scene. J. Neurophysiol., 88:438-454.

Diederich, A., 1992. Probability inequalities for testing separate activation models of divided attention. Perc. Psychophysics, 52:714-716.

Diederich, A., 1995. Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. J. Math. Psychol., 39:197-215.

Diederich, A., and H. Colonius, 1987. Intersensory facilitation in the motor component? Psychol. Res., 49:23-29

Diederich, A., and H. Colonius, 1991. A further test of the superposition model for the redundant-signals effect in bimodal detection. Perc. Psychophysics, 50:83-86.

Diederich, A., H. Colonius, D. Bockhorst, and S. Tabeling, 2003. Visual-tactile spatial interaction in saccade generation. Exp. Brain. Res., 148:328-337,

- Donders, F. C., 1868/1969. On the speed of mental processes. Acta Psychol., 30:412-431. (Translated by W. G. Koster.)
- Frens, M. A., A. J. Van Opstal, and R. F. Van der Willigen, 1995. Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Perc. Psychophysics, 57:802-816.
- Gielen, S. C. A. M., R. A. Schmidt, and P. Y. M. Van den Heuvel, 1983. On the nature of intersensory facilitation of reaction time. Perc. Psychophysics, 34:161-168.
- Driver, J., and C. Spence, 2000. Beyond modularity and convergence. Curr. Biol., 10:731-735.
- Hershenson, M., 1962. Reaction time as a measure of intersensory facilitation. J. Exp. Psychol., 63:289-293.
- Laurienti, P. J., J. H. Burdette, M. T. Wallace, Y. Yen, A. S. Field, and B.E. Stein, 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cog. Neurosci., 14:420-429.
- Lovelace, C. T., B. E. Stein, and M. T. Wallace, 2002. Now you hear it, now you don't: A light improves detection of a sound. Poster presented at Neural Control of Movement Conference, Naples, FL.
- Luce, R. D., 1986. Response Times: Their Role in Inferring Elementary Mental Organization, New York: Oxford University Press.

- Macaluso, E., C. D. Frith, and J. Driver, 2000. Modulation of human visual cortex by crossmodal spatial attention. Science, 289:1206-1208.
- Meredith, M. A., 2002. On the neural basis for multisensory convergence: a brief overview. Cog. Brain Res., 14:31-40.
- Meredith, M. A., and B. E. Stein, 1986. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol., 56:640-662.
- Meredith, M. A., J. W. Nemitz, and B. E. Stein, 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J. Neurosci., 10:3215-3229.
- Miller, J. O., 1982. Divided attention: evidence for coactivation with redundant signals. Cog. Psychol., 14:247-279.
- Molholm, S., W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe, 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cog. Brain Res., 14:115-128.
- Munoz, D. P., and R. H. Wurtz, 1995. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. J. Neurophysiol., 73:2313-2333.
- Nickerson, R. S., 1973. Intersensory facilitation of reaction time: energy summation or preparation enhancement. Psychol. Rev., 80:489-509.

Raab, D. H., 1962. Statistical facilitation of simple reaction times. Trans. New York Acad. Sci., 24:574-590.

Ricciard, L. M., 1977. Diffusion processes and related topics in biology, Berlin: Springer-Verlag.

Schwarz, W., 1989. A new model to explain the redundant-signal effect. Perc. Psychophys., 46:498-500.

Stein, B. E., and M. A. Meredith, 1993. The merging of the senses., Cambridge: MIT Press.

Todd, J. W., 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.

Tuckwell, H. C., 1988. Elementary applications of probability theory, London: Chapman and Hall.

Van Opstal, A. J., and D. P. Munoz, (this volume). Auditory-visual interactions subserving primate gaze orienting.

Van Zandt, T., 2002. Analysis of response time distributions, in Stevens' Handbook of Experimental Psychology (3rd Edition), Volume 4: Methodology in Experimental Psychology, (J. T. Wixted, vol. ed., H. Pashler, series ed.), New York: Wiley Press, pp. 461-516.

- Townsend, J. T., and F. G. Ashby, 1983. Stochastic modeling of elementary psychological processes, Cambridge, UK: Cambridge University Press.
- Townsend, J. T., and G. Nozawa, 1995. Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. J. Math. Psychol., 39:321-359.
- Ulrich, R., and M. Giray, 1986. Separate-activation models with variable base times: Testability and checking of cross-channel dependency. Perc. Psychophys., 34:248-254.
- Vroomen, J., and B. de Gelder, (this volume). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon.
- Welch, R. B., and D. H. Warren, 1986. Intersensory interactions, in Handbook of Perception and Human Performance, Vol.1: Sensory Processes and Perception, (K. R. Boff, L. Kaufmann, and J. P. Thomas, eds.), New York: Wiley, pp. 25-1 – 25-36.
- Worsley, K. J., 1981. An improved Bonferoni inequality and applications. Biometrika, 69:297-302.

Figure Captions

Figure 1: Panels on the left depict distributions f_A and f_V ; on the right the distributions of minima, f_{VA} . At the top panel f_A and f_V are fairly separated, i.e., they barely overlap. Thus, the mean for f_{VA} , on the right is about the same as for f_A . As the overlap between f_A and f_V increases, the distribution of minima shifts left and becomes most pronounced in the bottom panels where f_A and f_V overlap completely.

Figure 2: Testing separate activation in a visual-auditory focused attention task: area above the line indicates the degree of violation of Ineq. 4. Auditory intensity decreases from left to right, visual-auditory distance increases from top to bottom.

Figure 3:Figure depicts the trimodal RT data (stars) together with (1) the upper bound of Ineq. 7 obtained from the sum of two bimodal RT data minus the unimodal RT data (circles) and (2) the upper bound of Ineq. 5 obtained from the sum of three unimodal RT data (diamonds). Separate activation would predict an ordering of the curves reverse to what is seen (up to about 160 ms).

Figure 4: Superposition of visual, auditory, and tactile counter.

Figure 5 presents the fit of the model to data from a redundant target experiment with trimodal and bimodal stimuli including various stimulus onset asynchrony (SOA) conditions (Diederich, 1992).

Figure 6: The upper panels illustrate activation processes for different rates (drifts): the left hand side the one with the smaller drift representing a stimulus with weaker intensity. The lower panels demonstrate the processes for multimodal stimuli with different onset times.

Figure 7: The data from one of the subjects, saccadic reaction time to a visual target presented randomly left or right from fixation was reduced by up to 22% when accompanied by a spatially non-predictive tactile stimulus. This facilitation effect was larger for spatially aligned visual-tactile stimulus configurations than for contralateral presentation, and it increased with stimulus eccentricity (20° vs. 70°).

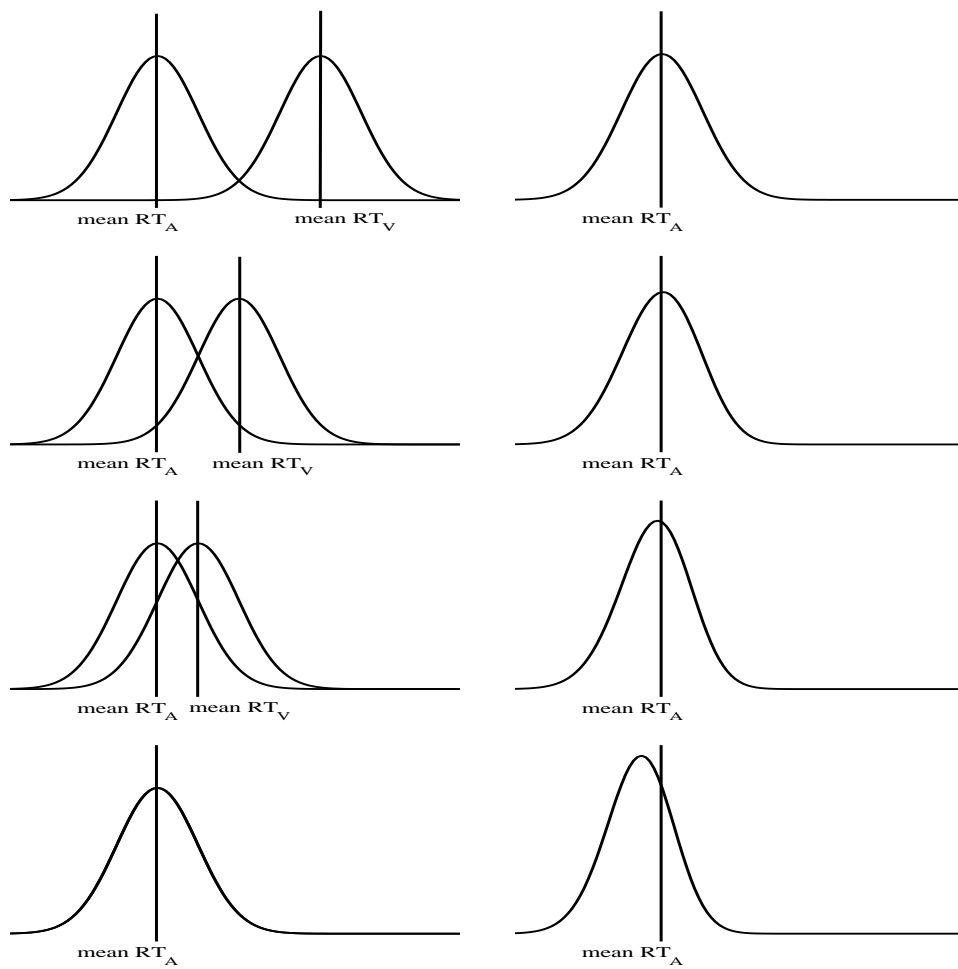


Figure 1:

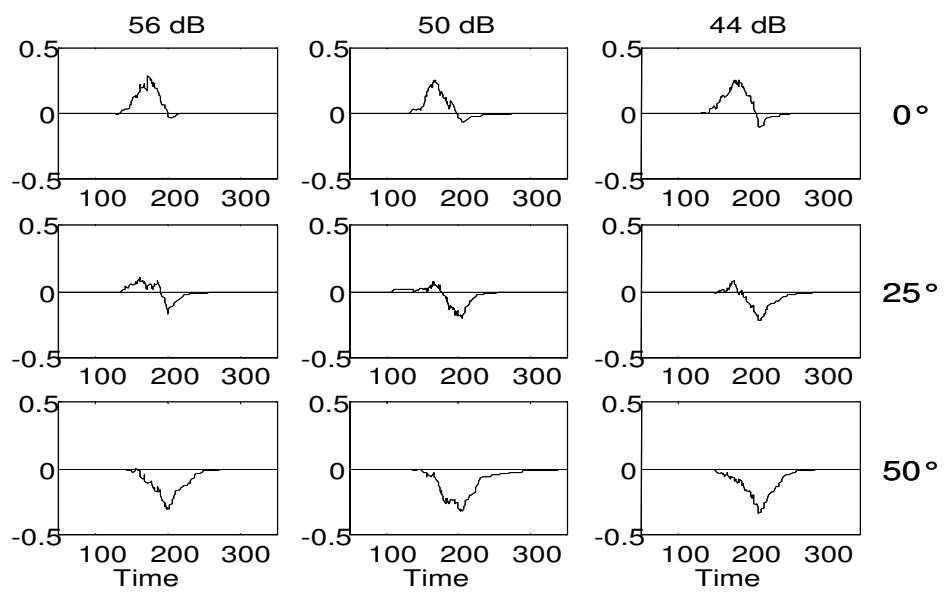


Figure 2:

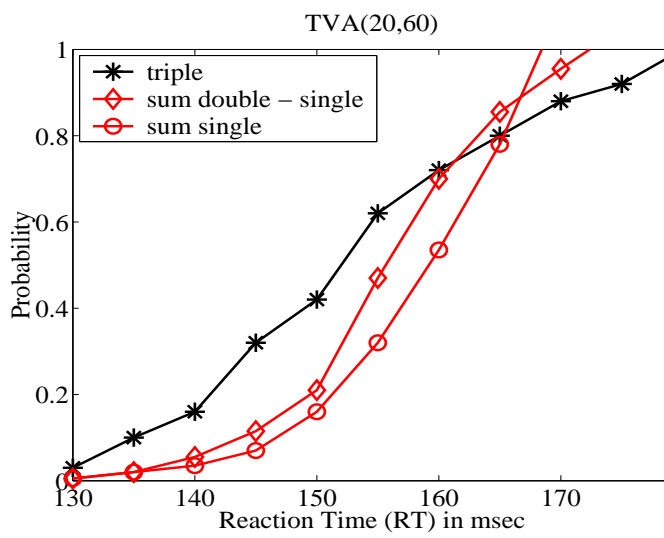


Figure 3:

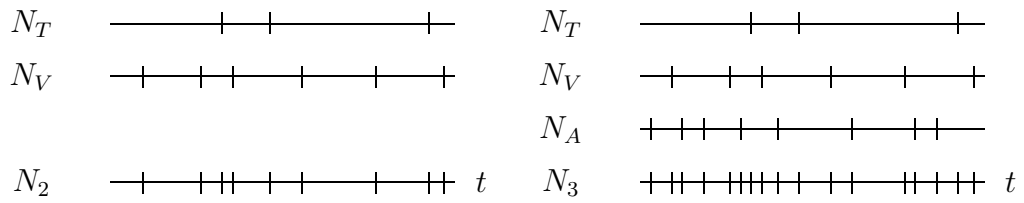


Figure 4:

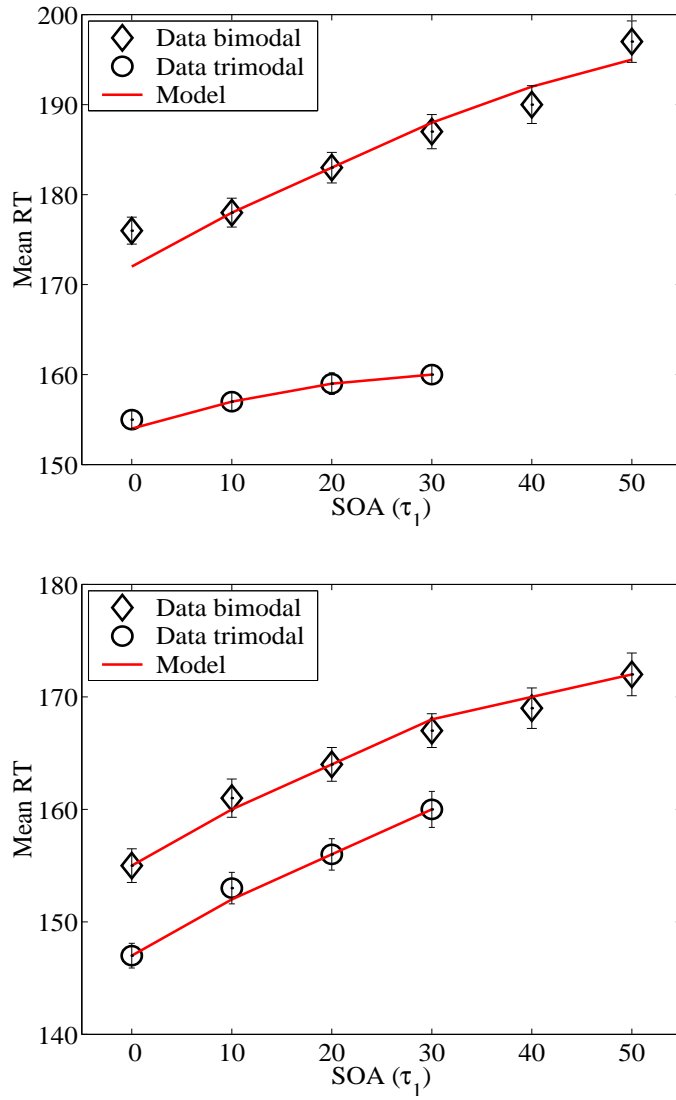


Figure 5:

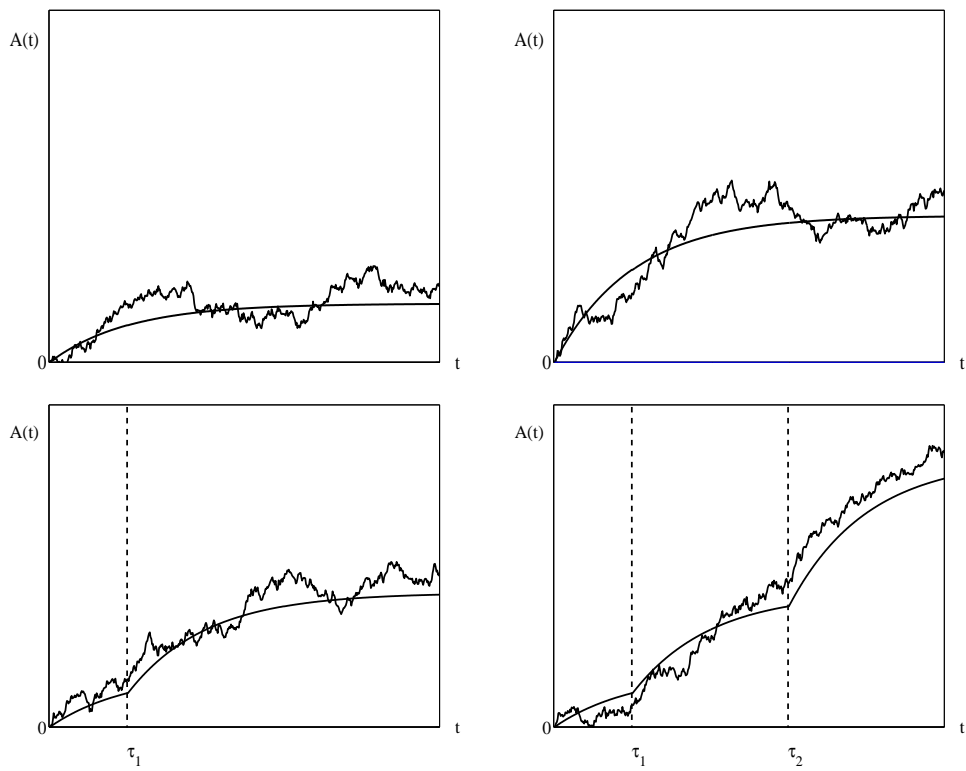


Figure 6:

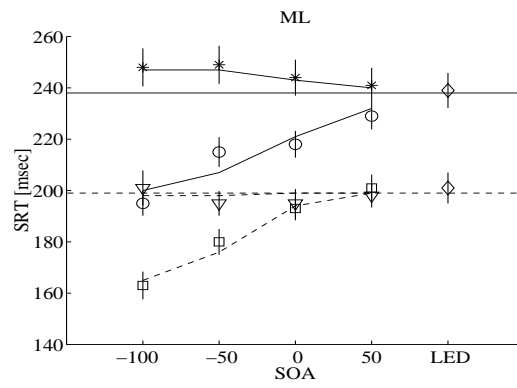


Figure 7: