Exp Brain Res
DOI 10.1007/s00221-006-0768-0

RESEARCH ARTICLE

Why two "Distractors" are better than one: modeling the effect of non-target auditory and tactile stimuli on visual saccadic reaction time

Adele Diederich · Hans Colonius

Received: 19 April 2006 / Accepted: 13 October 2006

Abstract Saccadic reaction time (SRT) was measured in a focused attention task with a visual target stimulus (LED) and auditory (white noise burst) and tactile (vibration applied to palm) stimuli presented as non-targets at five different onset times (SOAs) with respect to the target. Mean SRT was reduced (i) when the number of non-targets was increased and (ii) when target and non-targets were all presented in the same hemifield; (iii) this facilitation first increases and then decreases as the time point of presenting the non-targets is shifted from early to late relative to the target presentation. These results are consistent with the time-window-of-integration (TWIN) model (Colonius and Diederich in J Cogn Neurosci 16:1000–1009, 2004) which distinguishes a peripheral stage of independent sensory channels racing against each other from a second stage of neural integration of the input and preparation of an oculomotor response. Cross-modal interaction manifests itself in an increase or decrease of second stage processing time. For the first time, without making specific distributional assumptions on the processing times, TWIN is shown to yield numerical estimates for the facilitative effects of the number of non-targets and of the spatial configuration of target and non-targets. More generally, the TWIN model framework suggests that multisensory integration is a function of unimodal stimulus properties, like intensity, in the first stage and of cross-modal stimulus properties, like spatial disparity, in the second stage.

A. Diederich (✉)
School of Humanities and Social Sciences,
International University Bremen,
P.O. Box 750 561, 28725 Bremen, Germany
e-mail: a.diederich@iu-bremen.de

H. Colonius
Department of Psychology, University of Oldenburg,
P.O. Box 2503, 26111 Oldenburg, Germany
e-mail: hans.colonius@uni-oldenburg.de

Introduction

In a crossmodal *divided attention* task, stimuli from different modalities—typically, vision, audition, and touch—are presented in various modality combinations, either simultaneously or with a brief stimulus onset asynchrony (SOA). Participants are instructed to respond, often by manual key press, as soon as they detect any stimulus. The typical finding, already reported in Hershenson (1962), is a speed-up of responses to crossmodal stimuli relative to unimodal stimuli, and it is often referred to as *redundant target effect*. Note, that the term redundant target effect is also used for a similar finding when several targets from a single modality are presented. Many theoretical and experimental studies have since then been conducted to determine the causes of this effect (for a recent review, Diederich and Colonius 2004). The classic explanation for the redundant target effect is to assume that observed reaction time (RT) is the winner's time of parallel processes triggered by the visual, auditory, and tactile signals causing a *statistical facilitation* effect "ace model"; see Raab 1962). However, using Miller's inequality (Colonius 1990; Colonius and Diederich 2006; Diederich 1992; Miller 1982, 1986) as a benchmark test, responses to crossmodal stimuli have often been found to be faster than predicted by statistical facilitation (e.g., Patching and Quinlan 2004), leading to the development of an alternative account for the redundant target effect that assumes some coactivation mechanism to underlie the speed-up of responses (e.g., Diederich 1995; Schwarz 1994; Mordkoff and Yantis 1991; Townsend and Nozawa 1995; Townsend and Wenger 2004).

The focus of this paper is the study of crossmodal interaction effects occurring in a variant of the divided attention task where one of the modalities is singled out as "target" modality and participants are instructed to react as soon as they detect a target stimulus, while stimuli from the non-target modalities should be ignored (see Todd 1912). We prefer the neutral term "non-target" here to the more common terms "distractor" or "accessory stimulus" because this terminology presumes neither the existence of an effect nor its direction a-priori. In many experimental studies using this *focused attention* paradigm, vision is defined as target modality and often saccadic reaction times are collected to measure effects of crossmodal stimulation. In analogy to redundant target effect, a speed-up of saccadic reaction time (SRT) has been observed in focused attention experiments but, depending on both the spatial and temporal arrangement of target and non-target stimuli, inhibition or absence of any effect of non-targets on SRT have been reported as well (Amlôt et al. 2003; Colonius and Arndt 2001; Corneil and Munoz 1996; Corneil et al. 2002; Diederich et al. 2003; Frens et al. 1995; Harrington and Peck 1998; Hughes et al. 1998). Results also depend on details of the task, for example, the inclusion of "catch" trials where no target stimuli are presented, may result in certain subject strategies that are not considered in the following.

Given the different roles played by target and non-target stimuli in the focused attention task, the race model explanation suggested for the redundant target effect seems problematic in the case of the focused attention paradigm. Moreover, it provides no mechanism for the decrease in facilitation observed with increasing spatial disparity between the target and non-target stimuli. In the following, we outline a recently proposed model for crossmodal interaction effects in SRT (Colonius and Diederich 2004). This model builds on an explicit formalization of the important notion of a "time window of multisensory integration". Without making any specific parametric assumptions about probability distributions of the component processing times, the model nonetheless allows the numerical estimation of the probability of an intersensory interaction to occur in any given trial and, separately, of the average amount of facilitation taking place. After describing the model and its predictions, we report on an experiment testing important aspects of this modeling approach.

Time-window-of-integration (TWIN) model

Model assumptions

The initial anatomical separation of the afferent pathways for the different sensory modalities suggests that one can distinguish at least two serial stages of 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 some additional weak assumptions, a number of empirically testable predictions can be derived from this simple setup.

(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 an oculomotor 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. The following version of the third assumption is geared to the focused attention situation.

(3) Time-window-of-integration assumption (focused attention paradigm) Crossmodal interaction occurs only if (i) a non-target stimulus wins the race in the first stage, opening a "time window" such that (ii) the termination of the target peripheral process falls in the window. The duration of the "time window" is a constant. (Assuming a random duration of the time window would be a more realistic assumption, but using a constant of about 200 m has turned out to yield a satisfactory fit for most data sets so far.)

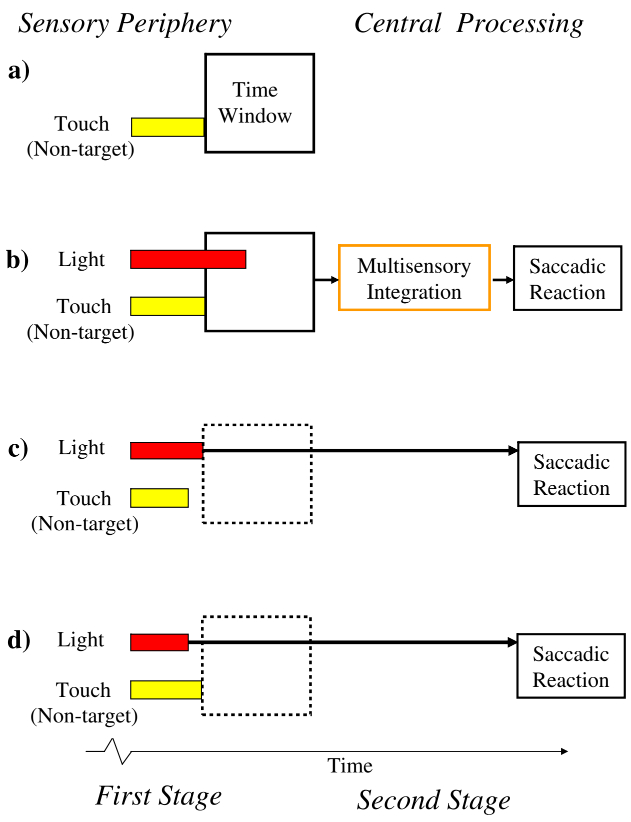

Fig. 1 Two-stage schema for TWIN model. Multisensory integration occurs only if peripheral processes terminate within the temporal integration window (see text). **a** The non-target, for example a tactile stimulus, opens the time window of integration, **b** the visual target stimulus fall within the time window and multisensory integration occurs, **c** the visual target stimulus wins the race but can not open the time window and, therefore, no multisensory integration is possible, **d** the non-target opens the window but the visual stimulus arrives so late that the window is already closed and no integration occurs

It may seem odd to postulate such target/non-target specificity for the window mechanism at a stage of stimulus processing that seems too early to afford a target/non-target identification of the winner. However, the effect of the non-target of winning the race in the first, peripheral stage should be seen as an increase of response readiness (similar to a "preparation enhancement", see Nickerson 1973) that is necessary, though not sufficient, for triggering the saccadic response[1]. The winning non-target will accelerate the saccadic response to the upcoming target stimulus if it falls into the time window, whereas in the case of the target being the winner, no discernible effect is expected compared to the unimodal situation. Figure 1 illustrates these ideas.

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 non-target 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 (2006).

The window of integration acts as a filter determining whether the afferent information delivered from different sensory organs is registered close enough in time for crossmodal interaction to take place. Passing this filter is necessary for crossmodal interaction to occur. It is not a sufficient condition because interaction also depends on the spatial configuration of the stimuli. Rather than assuming the existence of a joint spatiotemporal window of integration permitting interaction to occur only for both spatially and temporally neighboring stimuli, the TWIN model allows for interaction to occur even for rather distant stimuli of different modalities, as long as they fall exactly within the time window. Note, that this arrangement affords the organism more flexibility in a complex environment. In particular, response depression may occur with nearly simultaneous, but distant, stimuli.

Model predictions

Due to its two-stage processing assumption, it is easy to derive (see Appendix) that the expected amount of crossmodal interaction (ECI) in the TWIN model, defined as expected SRT in unimodal (target) trials minus expected SRT in crossmodal trials, is the product of two components: the probability of interaction to occur ($P(I)$, say) and the amount of interaction (in ms) (Δ, say):

$$\mathrm{ECI} \equiv E[RT_{\mathrm{unimodal}}] - \mathrm{E}[RT_{\mathrm{crossmodal}}] = P(I)\Delta. \qquad (1)$$

This *product rule* allows a number of empirically testable predictions to be made which do not require any specific distributional assumptions about the random

---

[1] At the neural level this would correspond to a gradual inhibition of fixation neurons (in superior colliculus) and/or omnipause neurons (in midline pontine brain stem).

processing times in the first or second stage. ECI, a theoretical entity, is obviously not observable in an experiment, but its value can be estimated from the finite data samples. Since, by assumption Eq. 4, the amount of interaction, Δ, does not depend on SOA, the dependence of expected amount of crossmodal interaction, ECI, on SOA is determined by how the probability of interaction, $P(I)$, is modulated through SOA. Consider two extreme SOA conditions: when the non-target 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 are small. On the other hand, if it is presented very early (large negative SOA), it is likely to win the race and to open the window, but the window may close before the arrival of the target. Again, the probability of interaction, $P(I)$, is small. Therefore, the largest effects are expected for some mid-range SOA values.

Prediction 1 *The expected amount of crossmodal interaction*, ECI, *increases*[2] *and then decreases as a function of stimulus onset asynchrony*.

The second prediction follows from the modulation of $P(I)$ as well: given that a precondition for crossmodal interaction to occur is that non-target processing wins the race in the first stage, having two (or more) non-targets of different modalities partaking in the race should increase the probability of one of them winning, thus increasing the probability of opening the window of integration.

Prediction 2 *The expected amount of crossmodal interaction*, ECI, *increases with the number of non-targets presented*.

The third prediction also capitalizes on the independence of Δ from the SOA values (Assumption 4).

Prediction 3 *For a fixed spatial configuration of the target and the non-targets, the expected amount of crossmodal interaction*, ECI, *is either positive (“facilitation”) across all SOA values or it is negative (“inhibition”) across all SOA values*.

Note that these predictions can be embedded in a more general framework that is based on the distinction between intra- and crossmodal stimulus properties. *Crossmodal* properties are defined when stimuli of more than one modality are present, like spatial distance of target to non-target or similarity between stimuli of different modalities. *Intramodal* properties, on the other hand, refer to properties definable for a single stimulus, no matter whether this property is definable in all modalities (like intensity) or in only one modality (like color or pitch).

Intramodal properties can affect the outcome of the race in the first stage and, thereby, the probability of interaction to occur. Crossmodal properties may affect the amount of crossmodal interaction (Δ) occurring in the second stage. Note that crossmodal features cannot influence first stage processing time since the stimuli are yet being processed in separate pathways. Initial empirical evidence for these predictions has been found in Colonius and Diederich (2004) for visual-tactile stimulation and in Arndt and Colonius (2003) for visual-auditory stimulation.

In this context, as suggested by an anonymous reviewer, it may be instructive to consider the so-called “inverse effectiveness rule” according to which multisensory enhancement effects are more pronounced when stimulus intensities are relatively low. This rule has been observed at the level of superior colliculus neurons (Wallace et al. 1996), at the behavioral level (Bernstein et al. 1970; see also Rach and Diederich 2006; Bolognini et al. 2005), and most recently in early cortical areas in fMRI data (Kayser et al. 2005). While it is not clear whether the mechanism underlying this effect of intensity level is the same in these different levels of observation, inverse effectiveness is actually predicted by the TWIN framework: In a focused attention task, reducing the intensity of the target stimulus diminishes its chances of winning the peripheral race, thus increasing the probability of interaction. Reducing the intensity of stimuli of all modalities simultaneously, whether in a focused or a divided attention task, will usually increase the reaction times. Consequently, the reaction time distributions will become broader resulting in a larger crossmodal SRT effect due to statistical facilitation.

Experiment

In order to test the predictions outlined above, we studied the effect of auditory and tactile non-targets on saccadic reaction time to a visual target under various temporal and spatial configurations of the stimulus set in a focused attention task. Independent variables were (i) the non-targets presented (one auditory, one tactile, or both auditory and tactile non-targets), (ii) the SOA between non-target(s) and target (five levels), and (iii) the relative spatial position of target and non-targets (ipsilateral, contralateral, visual-only).

---

[2] Note that in this paper “decreasing” is always meant as “decreasing or being constant” and “increasing” as “increasing or being constant”.

Materials and methods

Participants

Six undergraduate students, aged 19–22, 3 female, served as paid voluntary participants. All had normal or corrected-to-normal vision and were right-handed (self-description). They were screened for their ability to follow the experimental instructions (proper fixation, few blinks during trial, saccades towards visual target). They gave their informed consent prior to their inclusion in the study. The experiment was conducted in accordance with the ethical standards described in the 1964 Declaration of Helsinki.

Apparatus and stimulus presentation

Two red light-emitting diodes (LED, 25 mA, 3.3 mcd) presented against a black background served as visual targets. They were placed 20° to the left and right of a central fixation point (fixation LED, red, 25 mA, 5.95 mcd). Tactile stimuli were vibrations applied to the palm (50 Hz, 1 V, 1–2 mm amplitude) generated by two silenced oscillation exciters (Mini-Shaker, Type 4810, Bruel and Klær) placed 20° to the left and right of the fixation LED. Visual and tactile stimuli were positioned at the top of a table (180 × 130 × 75 cm) with a recess to sit in (referred to as *vertex*. The fixation LED was 38.5 cm away from the lower edge of the table. Auditory stimuli were bursts of white noise (59 dbA) generated by two speakers (Canton Plus XS). The speakers were placed horizontal to the participants' ear level 20° to the left and right of the fixation LED. Visual, tactile and auditory stimuli were 50, 52, and 120 cm apart from the vertex, respectively. Stimulus presentation was controlled by a PC-multifunction card.

Experimental procedure

The experiment was carried out in a completely darkened room so that participants were unable to see their hands during the experiment. The participant sat on a chair placed in the vertex, the head fixed on a chin rest. Every session began with 10 min of dark adaptation during which the measurement system was adjusted and calibrated. During this phase the participants put their hands at the position used during the entire experimental block. Thus, the participants were aware of the hand position and, thus, the position of the tactile stimulus.

Each trial began with the appearance of the fixation point. After a variable fixation time (800–1,500 ms), the fixation LED disappeared and, simultaneously, the visual target stimulus was turned on (no gap). Participants were instructed to gaze at the visual target as quickly and as accurately as possible ignoring any auditory or tactile non-targets (focused attention paradigm). Depending on the particular condition, the visual target appeared alone or in combination with tactile and/or auditory non-targets in various spatial configurations (see below). The onset of non-targets was shifted relative to the visual target by a stimulus onset asynchrony of –250, –100, –50, 0, or 50 m (negative values mean that the non-target was presented before the target). In trimodal trials, both non-targets shared the same SOA. The visual stimuli were presented for 500 ms; the auditory and tactile stimuli were turned off together with the visual. Thus, their duration varied between 750 and 450 ms, depending on SOA. Stimulus presentation was followed by a break of 2,000 ms in complete darkness before the next trial began, indicated by the onset of the fixation LED. Presentations were completely randomized over all conditions. After extensive training (60–120 min), each participant completed 14 blocks of 246 trials spread over 2 weeks, each block lasting about 15 min, resulting in a total of 3,444 trials (84 trials per condition).

Data collection

Saccadic eye movements were recorded by an infrared video camera system (EyeLink II, SR Research) with a temporal resolution of 500 Hz and horizontal and vertical spatial resolution of 0.01°. Criteria for saccade detection on a trial by trial basis were velocity (> 35°/s) and acceleration ($> 9{,}500\ °/s^2$). The recorded eye movements from each trial were checked for proper fixation at the beginning of the trial, eye blinks, and correct detection of start and end point of the saccade. Saccades were screened for anticipation errors (SRT < 80 ms), misses (SRT > 500 ms), and accuracy: trials with saccade amplitudes deviating more than 15° from the position of the target were excluded from the analysis (less than 2.7% of all data).

Results

We defined three ANOVA factors as follows: *laterality* with levels ipsilateral (i), contralateral (c), and ipsi- and contralateral (i-c), *SOA* (levels: –250, –100, –50, 0, 50 ms), and *non-targets* with levels auditory (A), tactile (T), auditory-tactile (AT), auditory and tactile presented in different hemifields (A–T), (T–A), and none. Thus, we use the following notation: Visual target only

(V), visual-auditory, presented ipsilateral (VA) and contralateral (V–A), visual-tactile, presented ipsilateral (VT) and contralateral (V–T) visual-auditory-tactile, presented ipsilateral (VAT) and contralateral (V–AT) and the two conditions with spatially separate non-targets, VA–T and VT–A, meaning that one non-target was presented in the same hemifield as the target and the second in the opposite hemifield. We conducted a four-way ($6 \times 3 \times 5 \times 6$) ANOVA on mean SRTs with *subjects* as random factor.

All main effects were significant at the $p < 0.001$ level (*subjects*: $F(5,21.7) = 85.45$, MSE = 603,655; *SOA*: $F(4,20) = 43.56$, MSE = 448,296; *laterality*: $F(1,5) = 64.13$, MSE = 117,109; *non-targets*: $F(3,15) = 11.77$, MSE = 15,129). The two-way interactions were significant as well, at the $p \leq 0.001$ level (*SOA* × *laterality*: $F(4,20) = 7.22$, MSE = 7,082; *SOA* × *non-targets*: $F(12,60) = 6.61$, MSE = 8,414; *laterality* × *non-targets*: $F(2,10) = 8.56$, MSE = 9,904). Except for the interaction *subject* × *SOA* ($F(20,31.26) = 7.08$, MSE = 10,298, $p < 0.001$), none of the two- and three-way interactions involving the *subject* factor were significant. We therefore pooled the data from all six subjects.

Post-hoc Tukey tests gave the following:

(1) Mean SRT to a visual target was significantly ($p < 0.001$) shorter (in ms) in the presence of non-targets: for ipsilateral presentation (MD = 18.6, SE = 1.2); for contralateral (MD = 12.9, SE = 1.2); and ipsi-contralateral (MD = 17.5, SE = 1.2). Further mean SRTs were shorter when target and non-targets were presented ipsilateral rather than contralateral (MD = 5.7, SE = 0.43, $p < 0.001$) as well as when presented ipsi-contralateral rather than contralateral (MD = 4.6, SE = 0.48, $p < 0.001$). No mean SRT difference was observed for ipsilateral compared to ipsi-contralateral presentation (MD = 1.1, SE = 4.8, $p = 0.097$).

(2) Mean SRT to the visual target was reduced when one or two non-targets were presented ($p < 0.001$). The effect was larger for two non-targets than for a single non-target, and the reduction from one to two non-targets was significant as well ($P < 0.001$). Moreover, the auditory non-target had a greater effect on mean SRT reduction than the tactile non-target ($p < 0.001$). In particular, mean difference and standard error were for auditory (A) (MD = 16.3, SE = 1.22); for tactile (T) (MD = 13.6, SE = 1.2); and auditory-tactile (AT, A-T, T-A) non-targets (AT:MD = 17.4, SE = 1.2; A-T:MD = 18.6, SE = 1.3; T-A:MD = 16.4, SE = 1.3). There was no difference between the effects of the auditory-tactile (AT, A-T, T-A) non-targets.

Averaged over all spatial conditions, mean SRT monotonically increased with SOA, i.e., the later the non-targets were presented. Figure 2 illustrates this for the trimodal conditions, Figs. 3 and 4 for the ipsilateral and contralateral (bi- and trimodal) conditions, respectively. However, given the significant interaction *subject* × *SOA*, no conclusions can be drawn about the existence of this monotonicity across or within individual participants. A posthoc analysis (Tukey) confirmed significant mean differences between all pairs of SOA values and between the SOA values and the visual-only condition ($P < 0.001$) across all spatial configurations, with the exception of SOA = 50 and the visual-only condition.

Crossmodal mean saccadic reaction times were reduced up to 30 ms compared to unimodal SRT (180 ms). Note that the monotonic increase of SRT with SOA is in line with Prediction 1 postulating a decrease of expected crossmodal interaction (ECI) with larger SOA values. In line with Prediction 2, mean SRT decreased with increasing numbers of non-targets present. Figure 5 illustrates this effect separately for the ipsi- and contralateral conditions.

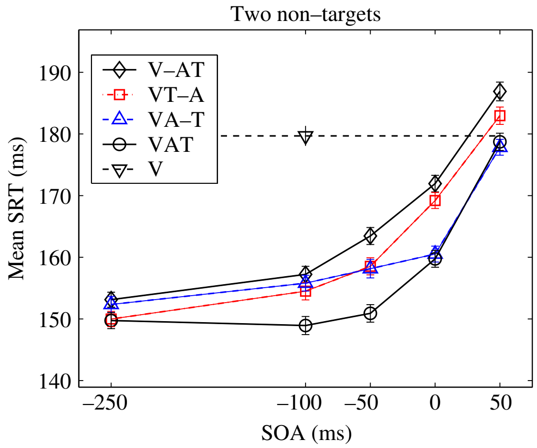

Fig. 2 Observed mean SRT (±SE) to trimodal stimuli as a function of SOA (relative to the target) and spatial configuration. The *dashed line* in the legend indicates contralateral configurations, e.g., V–AT means a trimodal stimulus condition where the auditory and tactile non-targets (AT) are presented on contralateral sides to the visual stimulus (V) (see also text). Observed unimodal SRT is presented by *dashed horizontal line*. The other curves refer to mean SRTs under different spatial configurations. Crossmodal facilitation takes place when these curves lie below the dashed horizontal line while crossmodal inhibition occurs when these curves lie above the dashed line. Each mean SRT consists of around 500 SRT (86 trials per condition, averaged across 6 participants, minus invalid trials)

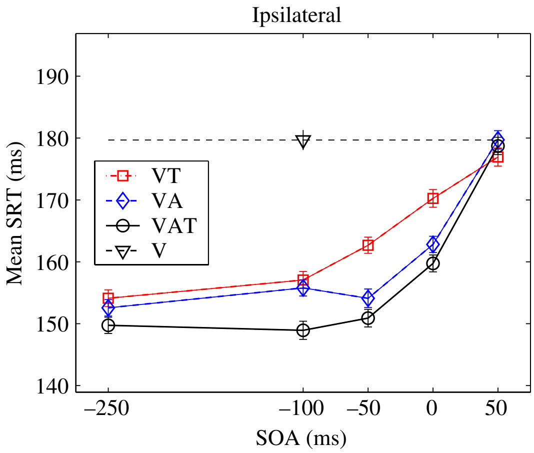

Fig. 3 Observed mean SRT (±SE) to ipsilateral bi- and trimodal stimuli as a function of SOA (relative to the target). Observed unimodal SRT is presented by dashed horizontal line. The other curves refer to mean SRTs under different spatial configurations. Crossmodal facilitation takes place when these curves lie below the *dashed horizontal line*. Each mean SRT consists of around 500 SRT (86 trials per condition, averaged across 6 participants, minus invalid trials)

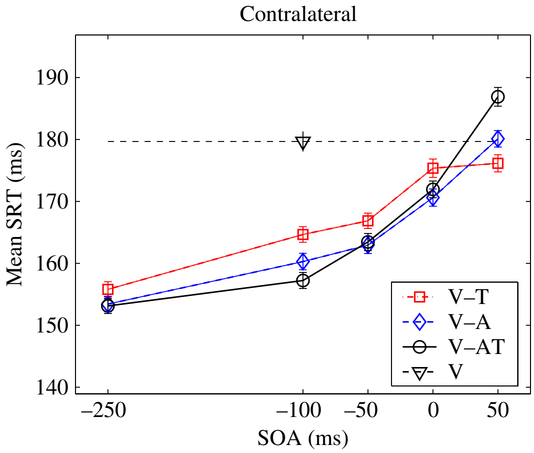

Fig. 4 Observed mean SRT (±SE) to contralateral bi- and trimodal stimuli as a function of SOA (relative to the target).The *dashed line* in the legend indicates contralateral configurations, e.g., V-AT means a trimodal stimulus condition where the auditory and tactile non-targets (AT) are presented on contralateral sides to the visual stimulus (V) (see also text). Observed unimodal SRT is presented by *dashed horizontal line*. The other curves refer to mean SRTs under different spatial configurations. Crossmodal facilitation takes place when these curves lie below the *dashed horizontal line* while crossmodal inhibition occurs when these curves lie above the *dashed line*. Each mean SRT consists of around 500 SRT (86 trials per condition, averaged across 6 participants, minus invalid trials)

Note that Prediction 3 requires that none of the graphs in Figs. 2, 3, 4 cross the horizontal line indicating the unimodal mean SRT because, otherwise, ECI would be positive for some SOA values and negative for others. The prediction holds true except for two configurations with contralateral auditory stimuli (i.e., V–AT and VT–A). Inspection of the individual data revealed that five out of our six participants do show this "late inhibition".

Estimating P(I) and Δ

The product rule allows an estimation of its factors, i.e., the probability of interaction $P(I)$ and the amount of interaction Δ, in the following relative sense. Let Δ^c and Δ^i denote the amount of interaction for contralateral and ipsilateral configurations, respectively (for simplicity, we assume a single Δ parameter for the contralateral conditions V–A, V–T, VT–A, VA–T, V–AT and another one for the ipsilateral conditions VT, VA VAT). Dividing the corresponding expected crossmodal interaction terms,

$$\frac{\mathrm{ECI}^c}{\mathrm{ECI}^i} = \frac{E[RT_{\text{unimodal}}] - E[RT^c_{\text{crossm.}}]}{E[RT_{\text{unimodal}}] - E[RT^i_{\text{crossm.}}]} = \frac{P(I)\Delta^c}{P(I)\Delta^i} = \frac{\Delta^c}{\Delta^i}. \tag{2}$$

Since this ratio does not depend on $P(I)$, it should be the same whether it is computed from conditions with one or with two non-targets. Inserting the corresponding sample mean SRT values yields an estimate of this ratio. In order to obtain a measure of its variability for statistical testing, we estimated the ratio for each participant separately (see Table 1). For example, for Participant 1, with a unimodal visual mean SRT of 155.2 ms, contralateral/one non-target mean SRT of 139.2 ms, and ipsilateral/one non-target mean SRT of 135.8 ms

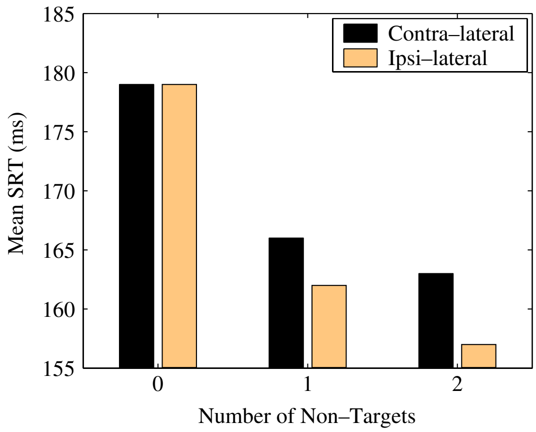

Fig. 5 Mean saccadic reaction time as a function of the number of non-targets across all conditions, separately for ipsi- and contralateral configurations (standard errors are too small to be visible)

Table 1 Participants' ECI ratios (mean crossmodal interaction for contralateral conditions estimating ECI^c, divided by mean crossmodal interaction for ipsilateral conditions estimating ECI^i) based on one-non-target configurations (fifth column) and on two-non-targets configurations (last column)

Participant	Unimodal RT_{uni}	One non-target			Two non-targets		
		RT_c	RT_i	$\mathrm{ECI}^c/\mathrm{ECI}^i$	RT_c	RT_i	$\mathrm{ECI}^c/\mathrm{ECI}^i$
1	155.17	139.27	135.76	0.82	135.55	131.63	0.84
2	180.67	168.06	164.51	0.79	166.15	157.85	0.65
3	190.39	183.43	179.76	0.63	183.31	177.39	0.65
4	168.36	158.30	151.17	0.58	154.85	146.64	0.68
5	186.37	172.65	169.31	0.79	168.00	162.32	0.81
6	197.69	178.00	175.03	0.87	173.37	169.78	0.91
Mean	179.69	166.73	162.71	0.75	163.71	157.74	0.76

See Eq. (2) in the text for the definition of ECIs. RT_{uni}, RT_c, and RT_i refer to mean SRT for unimodal, contralateral, and ipsilateral stimulus conditions, respectively

$$\frac{155.2-139.3}{155.2-135.8}=0.82;$$

analogously, for two non-targets,

$$\frac{155.2-135.5}{155.2-131.6}=0.84,$$

as found in the first row of Table 1. Given the small sample size ($N = 6$), the Wilcoxon signed rank test was conducted to test the null hypothesis of identical ratios in the one and two non-targets conditions. No evidence against the null was found ($P = 0.44$, 99%-confidence interval for difference: [–0.099 0.143]), but larger sample sizes with increased statistical power will be needed to bolster this conclusion. It should also be noted that the figures in Table 1 indicate that the amount of interaction in the contralateral configurations reaches between about 60 to about 90 percent of the interaction occurring in the ipsilateral configurations, depending on participant.

In a similar vein, estimates for the ratio of interaction probability with one versus with two non-targets, $P_1(I)/P_2(I)$ say, can be found by taking the ratio of the expected crossmodal interaction term based on 1–non-target configurations, $\mathrm{ECI}^{(1)}$, and the expected crossmodal interaction term based on 2–non-targets configurations, $\mathrm{ECI}^{(2)}$,

$$\frac{\mathrm{ECI}^{(1)}}{\mathrm{ECI}^{(2)}}=\frac{E[RT_{\mathrm{unimodal}}]-E[RT^{(1)}_{\mathrm{crossm.}}]}{E[RT_{\mathrm{unimodal}}]-E[RT^{(2)}_{\mathrm{crossm.}}]}=\frac{P_1(I)\Delta}{P_2(I)\Delta}=\frac{P_1(I)}{P_2(I)}. \quad (3)$$

Since $P(I)$ depends on SOA, these ratio estimates will vary with SOA as well. However, the Δ terms having vanished, the ratios should be identical whether computed for ipsi- or contralateral configurations. As before, this was tested statistically by considering the participants' data separately. The multivariate null hypothesis, that the interaction probability ratios for each value of SOA do not differ for ipsi- and contralateral configurations, could not be rejected (Hotelling's test, $F(4,7) = 1.015$, $P = 0.461$). The entries in the lower part of Table 2 (averages over participants) indicate that the probability of interaction to occur is smaller with one than with two non-targets, in line with Prediction 2. Moreover, there is a tendency of this difference to become more pronounced the later the non-targets are presented. Under these SOAs, the probability of a non-target winning the race should be relatively small so that the presentation of a second non-target will be relatively more beneficial for interaction to occur: a case of "statistical inverse effectiveness".

Discussion

The main experimental results reported here are in line with findings from previous experiments using visual targets and either auditory (e.g., Frens et al. 1995) or tactile (e.g., Amlôt et al. 2003) non-targets: (i) mean SRT is reduced in the presence of a non-target stimulus, (ii) the effect is the larger the earlier the non-target is presented, and (iii) the effect is most prominent when target and non-target are presented ipsilateral. Here we show, in addition, that the simultaneous presence of a non-target from a third modality (auditory or tactile) enlarges these effects resulting in a significant reduction of the saccadic reaction time with two non-targets relative to a single nontarget. Although auditory and tactile non-targets combine to reduce SRT, our data also suggest that the contribution of the auditory modality was stronger than that of the tactile in both ipsi- and contralateral configurations. In order to examine whether or not this is a genuine

Table 2 Participants' ratios of interaction probabilities $P_1(I)/P_2(I)$ (see Eq. 3 in the text) for contralateral (fifth column) and ipsilateral configurations (rightmost column) as a function of SOA values –250, –100, –50, and 0

Participant	SOA	Contralateral			Ipsilateral		
		$RT_{1\text{-non}}$	$RT_{2\text{-non}}$	$P_1(I)/P_2(\mathrm{I})$	$RT_{1\text{-non}}$	$RT_{2\text{-non}}$	$P_1(I)/P_2(\mathrm{I})$
1	–250	127.87	132.69	1.21	131.34	129.69	0.94
	–100	132.22	126.81	0.81	128.44	130.18	1.07
	–50	135.06	125.60	0.68	129.35	116.48	0.67
	0	147.90	137.64	0.41	137.32	130.54	0.72
2	–250	153.86	157.64	1.16	154.19	152.06	0.92
	–100	161.51	154.30	0.73	157.43	139.18	0.56
	–50	167.90	162.30	0.70	159.56	150.72	0.71
	0	175.89	172.47	0.58	171.20	163.47	0.55
3	–250	180.97	172.32	0.52	177.87	173.73	0.75
	–100	179.01	175.29	0.75	170.42	166.20	0.83
	–50	177.53	177.86	1.03	169.96	161.55	0.71
	0	187.12	184.45	0.55	184.74	180.47	0.57
4	–250	149.67	141.56	0.70	141.73	136.96	0.85
	–100	157.76	150.36	0.59	145.12	141.37	0.86
	–50	155.18	154.29	0.94	150.09	147.53	0.88
	0	162.83	157.21	0.50	153.26	143.96	0.62
5	–250	160.04	155.57	0.85	158.56	154.43	0.87
	–100	169.27	160.01	0.65	166.30	152.37	0.59
	–50	172.08	166.05	0.70	164.77	158.87	0.79
	0	177.97	171.38	0.56	172.29	165.32	0.67
6	–250	154.31	150.63	0.92	155.07	151.52	0.92
	–100	174.11	165.93	0.74	169.68	163.14	0.81
	–50	181.04	173.69	0.69	176.19	169.71	0.76
	0	186.10	179.89	0.64	180.25	173.88	0.73
Mean	–250	154.61	151.82	0.90	153.34	149.73	0.88
	–100	162.49	155.84	0.72	156.43	148.93	0.76
	–50	164.91	160.03	0.75	158.38	150.90	0.74
	0	173.00	167.22	0.54	166.52	159.74	0.66

The four bottom rows present averages over participants. RT_1–non and RT_2–non refer to mean SRT to the visual target when one and two non-targets are present, respectively

modality difference the experimental setup has to be extended in two aspects. First, the intensity levels of both non-targets should be varied systematically in order to rule out that the observed modality difference is merely due to the auditory non-target having been presented at a more salient level of intensity. Second, since both non-targets were always presented simultaneously, an additional onset asynchrony between the non-targets would have to be introduced in order to rule out that, due to different peripheral tactile and auditory processing times, the tactile modality was disadvantaged by not being presented at an optimal point in time.

The observed dependence of crossmodal interaction on SOA, spatial configuration, and number of non-targets was largely in accordance with the predictions from the TWIN model. Predicting a speed-up of SRT with the number of non-targets presented is arguably counterintuitive and its empirical confirmation lends further support to the model. There was, however, one violation of Prediction 3 concerning the slowing of crossmodal SRT in the 50 m SOA condition for some of the contralateral spatial configurations. This "late inhibition" of the saccadic response to the visual target by the non-target(s) has previously been observed in a similar experiment with auditory non-targets only (Colonius and Arndt 2001). It is not compatible with the basic TWIN assumptions because it would require a change from facilitation to inhibition (i.e., from positive to negative Δ) as a function of SOA. We have no solution to offer to this issue at present except to suggest that the effect is due to an additional mechanism not captured in the present version of the model. Note that the model also requires some extension with respect to the case when non-targets are presented much earlier than the limit of 250 ms employed here.

The notion that crossmodal integration is determined by a window of time had already been suggested in Meredith et al. (1987) recording from SC neurons, and now underlies many studies in this area (for a recent behavioral study, see Navarra 2005); see also Van Opstal and Munoz 2004. An important characteristic of TWIN is, in our view, that it lends itself to rigorous testing of the time-window notion due to its formal specification. Moreover, the fact that the TWIN model, without any assumptions about the statistical distributions of the underlying random processing times, allows an estimation of the relative probability of interaction ($P(I)$) and the amount of interaction (Δ) is another remarkable feature of the model. Although these parameters are not directly observable, hypotheses about how they depend on experimental variables such as SOA, intensity, spatial and temporal configuration, and the number of non-targets are directly testable via these estimates. Here these estimates revealed that (i) the addition of another non-target increased interaction probability $P(I)$ by 1/3 and that (ii) the amount of interaction Δ in contralateral configurations was only 77–79 percent of that in ipsilateral configurations. Importantly, the agreement of the estimates for these parameters being based on independent data sets from different experimental conditions is further evidence in favor of the model.

There are additional ways of testing the model. One that has been pursued in previous studies (Colonius and Diederich 2004; Diederich and Colonius 2004, 2006) is to postulate specific distributions for the random processing times in the first and second stages of TWIN, to estimate their parameters, and to make quantitative predictions for the intersensory interaction effects. Note that such an approach would also allow us to describe the differences among participants in detail. As indicated by the significant subject × SOA interaction, participants differ with respect to their general speed of reacting, and their idiosyncrasies can often be accounted for by individual parameter estimates, such as different widths of the temporal window implying different probabilities of interaction.

A recent study by Whitchurch and Takahashi (2006) collecting (head) saccadic reaction times in the barn owl lends further support to the notion of a race between early visual and auditory processes depending on the relative intensity levels of the stimuli. In particular, their data suggest that the faster modality initiates the saccade and the slower modality remains available to refine saccade trajectory. In addition, there is strong supportive evidence from a recent combined behavioral and neurophysiological study by Bell and colleagues (Bell et al. 2005) for the TWIN explanation of the ‘‘inverse effectiveness’’ rule mentioned earlier. They had monkeys make saccades to visual or audiovisual stimuli presented in spatial alignment or in opposite hemifields while simultaneously recording from neurons in the intermediate or deep layers of the superior colliculus (dSC). With low intensity, aligned audiovisual stimuli, the reduction in mean SRT was correlated with a reduction of the onset of (spike) activity preceding the visual burst triggering the saccade. With high intensity audiovisual stimuli, however, the onset of the visual response occurred so soon after stimulus presentation, ‘‘...leaving little opportunity for the auditory stimulus to bias the previsual activity before the arrival of the visual response in the dSC’’ (Bell et al. 2005, p. 3696). Thus, in terms of TWIN, since the auditory stimulus no longer has a ‘‘head-start’’ with high-intensity levels, the probability of the auditory stimulus opening a time window of integration, the probability of multisensory integration is strongly diminished. In fact, Bell and colleagues attribute the observed SRT reduction for high-intensity stimuli to increased premotor activity (activity after visual burst). It would certainly be interesting to extend the Bell et al. study to an experimental setup with varying audiovisual SOAs in order to test the TWIN predictions in more detail.

Finally, a word of caution about the realm of crossmodal effects covered by TWIN seems in order. It has repeatedly been proposed (e.g., Driver and Spence 1998; Macaluso and Driver 2001) that integration of spatial representations between sensory modalities does not rely solely on sensory convergence to multimodal areas but that it can also involve crossmodal influences upon sensory-specific cortices. In a recent fMRI study by Macaluso et al. (2005), subjects received visuo-tactile stimulation, spatially aligned or in different hemifields. Activity in visual extrastriate (lateral occipital) areas and in somatosensory parietal operculum was modulated by spatial congruence of the crossmodal stimulation, with stronger activations when concurrent visual and tactile stimuli were both delivered at the same contralateral location. Importantly, however, this difference in activity level occurred irrespective of which modality was task relevant and also of whether the stimuli were used to guide eye-movements or were just passively received. Macaluso et al. suggest that these effects reflect an automatic, stimulus-driven mechanism. In any event, given the independence of these effects from which modality was task-relevant and from whether or not saccades were required at all, it is obvious that such crossmodal effects on sensory-specific brain areas are outside of TWIN's sphere of validity.

Acknowledgment This research was supported by grants from Deutsche Forschungsgemeinschaft Di 506/8–1 and /–3.

Appendix

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

$$RT_{\text{crossmodal}} = W_1 + W_2, \tag{4}$$

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

$$\begin{aligned} &E[RT_{\text{crossmodal}}] \\ &\quad = E[W_1] + E[W_2] \\ &\quad = E[W_1] + P(I)E[W_2|I] + (1 - P(I))E[W_2|\text{not} - I] \\ &\quad = E[W_1] + E[W_2|\text{not} - I] \\ &\qquad - P(I)(E[W_2|\text{not} - I] - E[W_2|I]), \end{aligned}$$

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

$$E[RT_{\text{crossmodal}}] = E[W_1] + E[W_2|\text{not} - I] - P(I)\Delta. \qquad (5)$$

The term $P(I)\cdot\Delta$ can be interpreted as a measure of the expected crossmodal interaction effect in the second stage with positive Δ values corresponding to facilitation, negative ones to inhibition. In the unimodal condition, no interaction is possible. Thus,

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

and we arrive at the simple product rule for crossmodal interaction (CI)

$$ECI \equiv E[RT_{\text{unimodal}}] - E[RT_{\text{crossmodal}}] = P(I)\Delta \qquad (6)$$

discussed above in the section on model prediction. Note that Eq. (1) should hold for each crossmodal condition, with $P(I)$ depending on the number and type of target and non-target stimuli and Δ being a function of the specific spatial configuration.

References

Amlôt R, Walker R, Driver J, Spence C (2003) Multimodal visual-somatosensory integration in saccade generation. Neuropsychologia 41:1–15

Arndt A, Colonius H (2003) Two separate stages in crossmodal saccadic integration: evidence from varying intensity of an auditory accessory stimulus. Exp Brain Res 150:417–426

Bell AH, Meredith A, Van Opstal AJ, Munoz DP (2005) Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. J Neurophysiol 93:3659–3673

Bernstein IH, Rose R, Ashe VM (1970) Energy integration in intersensory facilitation. J Exp Psychol 86:196–203

Bolognini N, Frassinetti F, Serino A, Làdavas E (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp Brain Res 160:273–282

Colonius H (1990) Possibly dependent probability summation of reaction time. J Math Psychol 34:253–275

Colonius H, Arndt P (2001) A two-stage model for visual-auditory interaction in saccadic latencies. Percept Psychophys 63:126–147

Colonius H, Diederich A (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. J Cogn Neurosci 16:1000–1009

Colonius H, Diederich A (2006) The race model inequality: interpreting a geometric measure of the amount of violation. Psychol Rev 113:148–154

Corneil BD, Munoz DP (1996) The influence of auditory and visual distractors on human orienting gaze shifts. J Neurosci 16:8193–8207

Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. J Neurophysiol 88:438–454

Diederich A (1992) Probability inequalities for testing separate activation models of divided attention. Percept Psychophys 52(6):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, Colonius H (2004) Modeling the time course of multisensory interaction in manual and saccadic responses. In: Calvert G, Spence C, Stein BE (eds) Handbook of multisensory processes. MIT Press, Cambridge

Diederich A, Colonius H (2006) Modeling spatial effects in visual-tactile saccadic reaction time. Percept Psychophys (in press)

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

Driver J, Spence C (1998) Attention and the crossmodal construction of space. Trends Cogn Sci 2:254–262

Frens MA, Van Opstal AJ, Van der Willigen RF (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Percept Psychophys 57:802–816

Harrington LK, Peck CK (1998) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. Exp Brain Res 122:247–252

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289–293

Hughes HC, Nelson MD, Aronchick DM (1998) Spatial characteristics of visual-auditory summation in human saccades. Vison Res 38:3955–3963

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48:373–384

Macaluso E, Driver J (2001) Spatial attention and crossmodal interactions between vision and touch. Neuropsychology 39:1304–1316

Macaluso E, Frith CD, Driver J (2005) Multisensory stimulation with or without saccades: fMRI evidence for crossmodal effects on sensory-specific cortices that reflect multisensory location-congruence rather than task-relevance. Neuroimage 26:414–425

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci 10:3215–3229

Miller JO (1982) Divided attention: evidence for coactivation with redundant signals. Cogn Psychol 14:247–279

Miller JO (1986) Time course of coactivation in bimodal divided attention. Percept Psychophys 40:331–343

Mordkoff JT, Yantis S (1991) An interactive race model of divided attention. J Exp Psychol Hum Percept Perform 17:520–538

Navarra J, Vatakis A, Zampini M, Soto-Faraco S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. Cogn Brain Res 25:499–507

Nickerson RS (1973) Intersensory facilitation of reaction time: energy summation or preparation enhancement. Psychol Rev 80:489–509

Patching GR, Quinlan PT (2004) Cross-modal integration of simple auditory and visual events. Percept Psychophys 66:131–140

Raab DH (1962) Statistical facilitation of simple reaction times. Trans NY Acad Sci 24:574–590

Rach S, Diederich A (2006) Visual-tactile integration: Does stimulus duration influence the relative amount of response enhancement? Exp Brain Res 173:514–520

Schwarz W (1994) Diffusion, superposition, and the redundant-targets effect. J Math Psychol 38:504–520

Todd JW (1912) Reaction to multiple stimuli. In: Woodworth RS (ed) Archives of psychology no. 25. The Science Press, New York

Townsend JT, Nozawa G (1995) Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. J Math Psychol 39:321–359

Townsend JT, Wenger MJ (2004) A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. Psychol Rev 111:1003-1035

Van Opstal AJ, Munoz DP (2004) Auditory-visual interactions subserving primate gaze orienting. In: Calvert G, Spence C, Stein BE (eds) Handbook of multisensory processes. MIT Press, Cambridge, pp 373–393

Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. J Neurophysiol 76:1246–1266

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