

Time-Window-of-Integration (TWIN) Model for Saccadic Reaction Time: Effect of Auditory Masker Level on Visual–Auditory Spatial Interaction in Elevation

Hans Colonius · Adele Diederich · Rike Steenken

Received: 31 January 2009 / Accepted: 19 March 2009
© Springer Science+Business Media, LLC 2009

Abstract Saccadic reaction time (SRT) to a visual target tends to be shorter when auditory stimuli are presented in close temporal and spatial proximity, even when subjects are instructed to ignore the auditory non-target (focused attention paradigm). Previous studies using pairs of visual and auditory stimuli differing in both azimuth and vertical position suggest that the amount of SRT facilitation decreases not with the physical but with the *perceivable* distance between visual target and auditory non-target. Steenken et al. (Brain Res 1220:150–156, 2008) presented an additional white-noise masker background of three seconds duration. Increasing the masker level had a diametrical effect on SRTs in spatially coincident versus disparate stimulus configurations: saccadic responses to coincident visual–auditory stimuli are slowed down, whereas saccadic responses to disparate stimuli are speeded up. Here we show that the time-window-of-integration model accounts for this observation by variation of a perceivable-distance parameter in the second stage of the model whose value does not depend on stimulus onset asynchrony between target and non-target.

Keywords Multisensory integration · Saccadic reaction time · Auditory masker

Introduction

Saccades are fast and voluntary movements of the eyes made to align the foveal region of the retina with the spatial position of the stimulus eliciting the eye movement (e.g. Munoz and Schall 2004). While saccades are typically made to visual stimuli, they can also be elicited by auditory or somatosensory stimulation. In the *focused attention paradigm* (Diederich and Colonius 2004), participants are typically instructed to make a saccade as quickly and as accurately as possible toward a visual target stimulus suddenly appearing at a random position off the fixation point and, simultaneously, to ignore any co-occurring stimuli from other modalities. A pervasive result is that saccadic reaction time (SRT) toward a visual target is reduced in the presence of a spatially or temporally aligned non-target auditory stimulus, and that this facilitatory effect diminishes, or even reverses into inhibition, when the spatiotemporal distance between the stimuli increases (e.g., Colonius and Arndt 2001; Corneil and Munoz 1996; Diederich and Colonius 2007a, b; Frens and van Opstal 1998; Harrington and Peck 1998; Hughes et al. 1994, 1998; Lueck et al. 1990; for reviews, see Diederich and Colonius 2004; and Van Opstal and Munoz 2004). Specifically, Frens et al. (1995) found that saccadic latencies increased with about 0.5 ms per degree stimulus separation, up to about 35 ms, using broadband white noise as auditory stimuli.

The fact that one can find such a lawful relationship suggests that the information about the magnitude of visual–auditory spatial disparity has mandatory access to the mechanism of saccade initiation. This seems remarkable given that participants are not at all required to localize the auditory stimulus in the focused attention paradigm: they are only asked to move the eyes toward the

This article is published as part of the Special Issue on Multisensory Integration.

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

A. Diederich
Jacobs University Bremen, Bremen, Germany

visual target ignoring any non-target auditory stimulation. In particular, the Frens et al. (1995) study revealed that it is not the physical but, rather, the *perceivable* distance between visual and auditory stimulus position that matters (Frens et al. 1995, pp. 807–808).

In order to further elucidate the nature of this multisensory integration mechanism, Steenken et al. (2008) systematically varied perceivable (vertical) distance between visual target and auditory non-target by adding auditory background noise, varying in intensity, to the visual–auditory stimulus combinations. The purpose of this paper is to demonstrate how the results from the Steenken et al. study can be accounted for within the framework of the time-window-of-integration (TWIN) model for multisensory integration in saccades (Colonus and Diederich 2004). First, we give some background including the neural underpinnings of the effects, then we summarize the results of Steenken et al. (2008) and subsequently present the TWIN framework and its interpretation of the data.

Background

Neural correlates of multisensory orienting behavior have been found primarily in the midbrain superior colliculus (SC). Neurons in the intermediate/deep layers of the SC (dSC) are involved in the initiation and control of saccades (for review, see Munoz and Fecteau 2002; Sparks 1999; Sparks et al. 2001). Information about stimulus location is represented topographically - the horizontal dimension is mapped rostrocaudally, the vertical dimension mediolaterally on the SC (Middlebrooks and Knudsen 1984)—by an arrangement of neurons according to the location of their receptive fields (RFs). Many of these same neurons exhibit multisensory activity paralleling the spatiotemporal rules found in behavioral studies (King and Palmer 1985; Meredith and Stein 1986; Populin and Yin 2002; Wallace et al. 1996). The spatial register between the auditory and visual sensory maps is formed by multisensory neurons whose different RFs are in register with one another yielding a common frame of reference (Stein and Meredith 1993). Their firing rates are greatest for spatially aligned stimuli and decrease in magnitude as spatial disparity increases (Bell et al. 2001; Frens and Van Opstal 1998; Meredith and Stein 1996). These sensory maps are also in register with the premotor maps found in SC (e.g., McIlwain 1986), and many SC neurons are involved in both sensory and motor maps. Recent studies by Bell and colleagues (Bell et al. 2005, 2006) demonstrated a particularly close link between changes in neural activity related to stimulus modality with changes in gazing behavior of alert monkeys.

When both stimuli originate from the same location in space, they are likely to fall within the respective excitatory

RFs of the same multisensory SC neuron, thereby triggering a superadditive response enhancement of the neuron's activity. For spatially disparate presentations, when the stimulus of one modality is falling outside the borders of the neuron's RF, response depression occurs as a result of the antagonism between an inhibitory input derived from activation of the extra-receptive field region of that stimulus and the excitatory input from the within-field stimulus of the other modality (Stein 1998; Kadunce et al. 1997). Note that the graded decline of response enhancement of SRT typically observed in behavioral studies is consistent with these RF mechanisms if one assumes that orienting behavior is the collective result of a potentially large number of multisensory neurons with gradually shifted RFs.

Nevertheless, the integration of multisensory signals into an orienting response is far from trivial since different sensory modalities are initially transduced separately and encoded in different frames of reference. The *oculocentric* frame of reference in which saccades are represented must be derived from retinotopic signals for visually guided saccades and from head-centered space for aurally guided saccades. This latter transformation is particularly complex because the head-centered space is constructed from different acoustic cues: Whereas the azimuth of a sound source is derived from binaural cues, such as interaural timing and intensity differences, estimating the elevation component is based on spectral filtering by the pinnae/head and, in the case of a tonal stimulus, this monaural cue cannot deliver unambiguous information on the vertical sound source position (Wightman and Kistler 1989; Blauert 1997). For example, Frens et al. (1995) showed that, with 700-Hz tones as auditory non-target, only the horizontal separation determined the amount of SRT reduction, the actual vertical position did not play a role.

Further information on the time course of SRT enhancement comes from a study by Heuermann and Colonus (2001). They presented visual–auditory stimulus pairs varying in both elevation and azimuth with stimulus onset asynchrony (SOA) ranging from –60 to 40 ms, where negative SOA values mean that the auditory non-target was presented prior to the visual target. As no maskers were presented, the auditory white-noise stimuli were easily localizable with maximal bimodal enhancement (i.e., bimodal SRT minus unimodal SRT) for the coincident condition at SOA = –60 ms. Interestingly, there was no difference in the level of enhancement between pairs differing in azimuth only and pairs differing in both elevation and azimuth, as long as the auditory stimulus was presented simultaneous to or after the visual stimulus; bimodal enhancement, however, was still observable under both conditions. Presumably, when the auditory stimulus was presented “too late” there was no

time for the elevation component to be computed by the pinnae/head system utilizing binaural cues—which, in our estimate, is some 30 ms slower than the horizontal system—so that saccade initiation was already under way. In other words, the *perceivable* distances between the visual and auditory stimulus at the time of saccade initiation were about the same in the two types of configuration.

Effect of Auditory Masker Level on Visual–Auditory Spatial Interaction (Steenken et al. 2008)

Following up on the hypothesis that *perceivable* distance between visual target and auditory non-target determines multisensory interaction in SRT, Steenken et al. (2008) manipulated perceivable distance by introducing an uncorrelated white noise masker of varying intensity level at the beginning of the trial and lasting for 3 s, in addition to the auditory non-target. Earlier, in an auditory localization task, Good and Gilkey (1996) had shown that with decreasing signal-to-noise ratio pointing responses in an up/down judgment task are less and less correlated with the target position, suggesting that variability of perceivable distances increases with increasing masker level. Since the localization of acoustical stimuli in elevation is already affected at higher signal-to-noise ratios compared to the azimuth (Good and Gilkey 1996; Lorenzi et al. 1999), Steenken et al. varied spatial positions of target and non-target in the vertical axis only (one position above and one below fixation point). In the *coincident* condition, visual target (red light-emitting diode) and auditory non-target (1/f noise, range 100–22 KHz) of 100 ms duration were both presented at the top or at the bottom position, whereas in the *disparate* condition the stimulus of one modality was presented at the top and the stimulus of the other modality was presented at the bottom position (for further detail, refer to the original article).

In a localization task, conducted in blocks of trials separate from the SRT measurement, Steenken et al. (2008) observed the expected decrease of auditory localizability from perfect to guessing level (measured in d -prime) with increasing masker intensity (4 levels including a no-masker condition). In the focused attention task, the auditory non-target was presented 60 or 30 ms before, simultaneous, or with a delay of 30 ms to the visual target.

The main findings for SRTs were as follows (see Fig. 1): (i) SRT in the coincident configuration was faster than in the disparate configuration, but that difference vanished with increasing masker level in such a way that saccade initiation to the common target/non-target position (top or bottom) was slowed down with increasing masker level; (ii) at the same time, saccade initiation to the target opposite to the non-target (disparate condition) was accelerated with

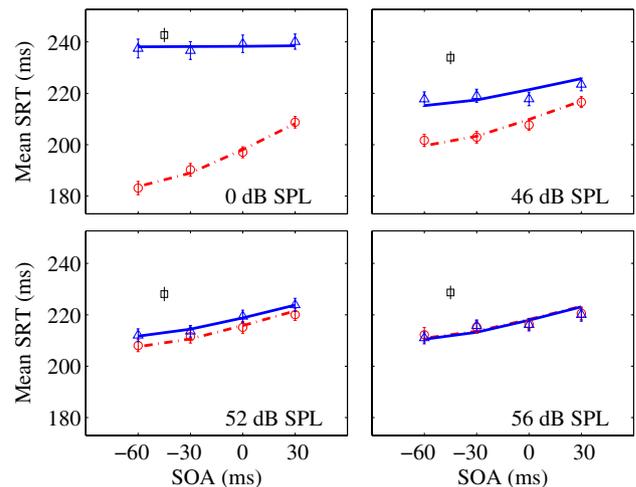


Fig. 1 Mean saccadic reaction time (SRT) as a function of stimulus onset asynchrony between visual target and auditory non-target for auditory masker levels 0, 46, 52, and 55 dB. The continuous (*dashed*) lines refer to predicted SRT in the disparate (coincident) configuration, respectively. Data points \pm one standard error are presented (averaged across participants). Unimodal visual mean SRT (indicated by the separate point above the lines) was not fitted in this version (see text)

masker level; (iii) in the coincident configuration, SRT was the faster the earlier the auditory non-target was presented; (iv) unimodal SRT to the visual target also decreased with masker level, but still remained slower than bimodal SRT.

The seemingly contradictory findings (i) and (ii) can be reconciled by postulating that the increased variability of perceivable distances, due to increasing masker level, has opposite effects under the two configuration conditions: in the coincident condition, the diminishing localizability of the auditory stimulus with increasing masker level should trigger the occurrence of larger and larger perceivable distances between visual and auditory stimulus even though both stimuli remain at their (nearly) identical physical position. By contrast, in the disparate condition the masker level increase should allow the occurrence of smaller and smaller perceivable distances between the stimuli even though their physical vertical distance remains invariant. Thus, increasing masker level (i.e., decreasing signal-to-noise ratio) will increase the average perceivable distance in the coincident condition and will decrease it in the disparate condition such that the average perceivable distances become identical in the limit. Let us call this the *masker level-distance hypothesis*.

Time-Window-of-Integration (TWIN) Model: Main Assumptions

Early models of crossmodal reaction time enhancement were commonly based on the notion of a parallel

independent race among separate sensory channels (Raab 1962; Miller 1982). Upon mounting evidence of enhancement effects being larger than predictable by this probability summation mechanism, a number of alternative models have been developed (Diederich and Colonius 2004, for a review). The TWIN model (Colonius and Diederich 2004) distinguishes 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). In the first stage, a race among the peripheral neural excitations in the visual and auditory pathways triggered by a crossmodal stimulus takes place. The second stage comprises neural integration of the input and preparation of an oculomotor response. Thus, the model retains the classic notion of a race mechanism as an explanation for crossmodal interaction but restricts it to the very first stage of stimulus processing. The second stage is defined by default: it includes all subsequent, possibly temporally overlapping, processes that are not part of the peripheral processes in the first stage. The TWIN model makes specific assumptions about the temporal configuration needed for multisensory integration to occur.

Time-Window-of-Integration Assumptions

In the focused attention paradigm, crossmodal interaction occurs only if (a) a non-target stimulus wins the race in the first stage, opening a “time window“ such that (b) the termination of the target peripheral process falls in the window; (c) the duration of the “time window” is a constant. The idea here is that the winning non-target will keep the saccadic system in a heightened state of crossmodal reactivity such that the upcoming target stimulus, if it falls into the time window, triggers crossmodal interaction. At the neural level, this might correspond to a gradual inhibition of fixation neurons (in SC) and or omnipause neurons (in midline pontine brain stem). In the case of the target being the winner, no discernible effect on saccadic reaction time is predicted, analogous to the unimodal situation.

Assumption of Spatio-Temporal Separability

The amount of interaction (facilitation or inhibition) 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). 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. Note that rather than assuming the existence of a joint spatio-temporal window of integration permitting interaction to occur only for both spatially and temporally neighboring stimuli, TWIN allows for interaction to occur even for rather distant stimuli of different modality, as long as they fall within the time window.

The above assumptions are part of a more general framework making a distinction between intra- and cross-modal 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 for all modalities (like intensity) or in only one modality (like shape). Intramodal properties can affect the outcome of the race in the first stage and, thereby, the probability of an interaction. Crossmodal properties may affect the amount of crossmodal interaction occurring in the second stage. Note that crossmodal features cannot influence first stage processing time since the stimuli are still being processed in separate pathways. Further elaboration of TWIN includes a mechanism to distinguish between the role of the non-target as a warning signal and its contribution to “true” multisensory integration (Diederich and Colonius 2007a, b; 2008a, b).

An important, testable prediction following from the assumptions is that the *expected* amount of crossmodal interaction (ECI), defined as difference between mean SRT in the unimodal visual condition and mean SRT in the bimodal condition, can be written as a product: the probability of integration $P(I)$, a factor depending on SOA and intramodal properties only, *times* the amount of interaction, Δ , a factor depending on crossmodal properties only, in particular the distance between target and non-target:

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

Δ takes on positive or negative values (or zero) depending on whether multisensory integration has a facilitative or inhibitory effect on SRT (see appendix for a more details on the model).

Time-Window-of-Integration (TWIN) Model: Effect of Masker

Although it obviously must undergo processing in the auditory periphery, it is assumed here that the masker does not play a role as a non-target stimulus modulating multisensory integration: the masker is switched on at the beginning of the trial, and visual target and auditory non-target stimuli only appear after a random foreperiod of 740–1500 ms after the masker, making it very unlikely for

Table 1 Ratios (coincident/disparate) of amount-of-interaction parameter (Δ) values [in ms] as a function of stimulus onset asynchrony (SOA) and auditory masker level

SOA (ms)	Masker Level (db)			
	0	46	52	55
-60	8.7	2.1	1.1	0.8
-30	7.9	2.0	1.1	1.0
0	11.5	1.6	1.1	1.2
30	7.0	1.8	1.8	1.25

the masker to fall into a common time window of integration with these stimuli. Nevertheless, an effect on unimodal visual SRTs of about 12 ms by increasing masker level was observed (see Fig. 1), reflecting a slight enhancement that is arguably due to some arousal in central processing or motor components.

In line with the assumptions of TWIN and following from the hypothesis of an influence of masker level on perceivable distance between target and non-target, any observed spatially specific effect of the masker on bimodal SRTs should be reflected in the amount-of-interaction parameter, Δ . Before presenting a parametric fit of TWIN to the data, it is informative to consider Eq. 1 again because it allows checking a prediction of TWIN without any specific distributional assumption. Writing $ECI^{(c)}$ and $ECI^{(d)}$ for the expected amount of crossmodal interaction in the coincident and disparate condition, respectively, Eq. 1 implies

$$\frac{ECI^{(c)}}{ECI^{(d)}} = \frac{P(I) \cdot \Delta^{(c)}}{P(I) \cdot \Delta^{(d)}} \quad (2)$$

for each level of the masker, with $\Delta^{(c)}$ and $\Delta^{(d)}$ the amount-of-interaction parameters in the coincident and disparate condition, respectively. The term on the left-hand side is estimated by inserting the corresponding observed mean SRT values. For the right-hand side, the probability of integration, $P(I)$, depends on SOA but not on spatial configuration and, therefore, cancels, whereas $\Delta^{(c)}$ and $\Delta^{(d)}$ do not depend on SOA. Therefore, the ratio should remain invariant (except for random variability) across all SOA values, with a separate value for each level of the masker. Moreover, under the masker level-distance hypothesis, the ratio should converge towards 1 with large enough masker levels. The result of this computation is presented in Table 1.

Taking into account that the computed cell entries are numerical ratios, variability within each column appears quite low and, arguably, this non-parametric test does not indicate evidence against the model.¹

¹ For a statistical treatment of this issue (confidence intervals), a non-parametric bootstrap method seems indicated and is planned in future work.

In order to derive further predictions from the TWIN model, we need to specify the probability distributions for the processing times in the first stage. For simplicity, we assume exponential distributions for the peripheral processing time V for a visual target and A for an auditory non-target, with parameters λ_V and λ_A , respectively. The expected response times for the crossmodal conditions and the unimodal (visual) condition then are

$$E[RT_{crossmodal}] = \frac{1}{\lambda_V} + \mu - P(I) \cdot \Delta \quad (3)$$

$$E[RT_{unimodal}] = \frac{1}{\lambda_V} + \mu, \quad (4)$$

respectively. Here, the mean of second stage processing time (without interaction occurring) is taken to be μ , where we need not specify the underlying distribution as long as predictions are restricted to the expected values of SRT.

For each combination of spatial configuration (coincident vs. disparate) times masker level (4 levels), an amount-of-interaction parameter (Δ) represents the effect of average perceivable distance. Note that under the masker level-distance hypothesis these parameters must be ordered: decreasing in the coincident configuration and increasing in the disparate configuration, with increasing masker levels. The final parameter to be estimated is size of the time window, ω .

Parameters were estimated by minimizing a Pearson χ^2 statistic (see appendix) and resulted in the values given in Table 2. As depicted in Fig. 1, the fit is nearly perfect; this is not really surprising given the large number of parameters (13) relative to the number of observations (32). More importantly, however, the parameter values follow a plausible pattern: (i) auditory peripheral processing (60 ms) is faster than visual (105 ms); (ii) the amount of interaction Δ in the coincident configuration decreases with increasing masker level; (iii) for the disparate configuration, there is no interaction in the absence of the masker ($\Delta_0^{(d)} = 1$ ms), and the interaction for the masker conditions slightly increases towards a level close to the highest-masker coincident configuration (28–34 ms). The unimodal visual mean SRT was not fitted here but, in a version of the fit not presented here, adding an extra parameter for expected second stage processing time in the unimodal condition, a similar fit was obtained.

Discussion

The central finding here is that the time window of integration model can account for the dependence of the crossmodal interaction effect on the level of an auditory masker. In Steenken et al. (2008), a diametrical effect on saccadic reaction time was observed with increasing

Table 2 Estimated parameter values of TWIN

$1/\lambda_V$	$1/\lambda_A$	μ	ω	$\Delta_0^{(c)}$	$\Delta_{46}^{(c)}$	$\Delta_{52}^{(c)}$	$\Delta_{55}^{(c)}$	$\Delta_0^{(d)}$	$\Delta_{46}^{(d)}$	$\Delta_{52}^{(d)}$	$\Delta_{55}^{(d)}$	χ^2
105	60	133	400	66	47	37	34	1	28	32	34	2.26

masker level: for a spatially coincident configuration of a visual target and an auditory non-target saccadic responses were slowed down, whereas responses were facilitated when target and non-target were presented at vertically opposite positions. As localizability of an auditory stimulus decreases with increasing masker level, it is hypothesized that the average *perceivable* distance between target and non-target is also affected in diametrical way: it increases for the coincident configuration and it decreases for the disparate configuration (*masker level-distance hypothesis*). Perceivable distance clearly being a crossmodal property in the TWIN framework, its effect on multisensory interaction should be captured by the amount-of-interaction parameter in the second stage of the model (Δ) which does not depend on the stimulus onset asynchrony of target and non-target. This was corroborated by a distribution-free empirical test based on Eq. 2. Moreover, the pattern of Δ values was in perfect accordance with the prediction from the masker level-distance hypothesis, while all other model parameters remained the same across masker levels. Note that the masker level-distance hypothesis is not part of the TWIN model framework proper but, rather, calls for a psychophysical model (or, at least, stochastic mechanism) of how localizability of an auditory non-target stimulus relates to its perceivable distance from a visual target.

The results presented here provide further support for the time window of integration model that has by now been tested by the authors in a variety of ways: it accounts for varying the spatial configuration of the stimuli (Diederich and Colonius 2007b), for the effect of increasing the number of non-targets presented together with the target (Diederich and Colonius 2007a), for the warning effect occurring with large negative SOAs (Diederich and Colonius 2008a), for the effects of increasing the intensity of the non-target (Diederich and Colonius, 2008), and for age effects on crossmodal interaction (Diederich et al. 2008). There is also mounting neurophysiological evidence for the TWIN postulate of separability of temporal from non-temporal aspects in multisensory integration (e.g., van Attefeldt et al. 2007; Bell et al. 2006; van Opstal and Munoz 2004).

It is interesting that the idea of multisensory integration being determined by a time window had already been suggested in a study by Meredith et al. (1987) recording from SC neurons. This notion, together with the concept of a race among peripheral processes, now underlies many studies of crossmodal temporal interaction (e.g., Lewald and Guski 2003; Morein-Zamir et al. 2003; Spence and Squire 2003; see Whitchurch and Takahashi 2006, for head

saccades in the barn owl). Finally, the importance of modeling the time course of multisensory integration is increasingly being recognized in recent studies (Rowland et al. 2007; Rowland and Stein 2007; see also Ma and Pouget 2008).

Acknowledgment This research was supported by Grants from Deutsche Forschungsgemeinschaft Di 506/8-1 and Di 506/8-3 and SFB-TR31 (Active Listening).

Appendix

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

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

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

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

$$RT_{crossmodal} = S_1 + S_2, \quad (5)$$

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

$$\begin{aligned} E[RT_{crossmodal}] &= E[S_1] + E[S_2] \\ &= E[S_1] + E[S_2|\text{not-}I] \\ &\quad - P(I)(E[S_2|\text{not-}I] - E[S_2|I]), \end{aligned}$$

where $E[S_2|I]$ and $E[S_2|\text{not-}I]$ denote the expected second stage processing time conditioned on interaction occurring (I) or not occurring ($\text{not-}I$), respectively. Setting

$$\Delta \equiv E[S_2|\text{not-}I] - E[S_2|I]$$

this becomes

$$E[RT_{crossmodal}] = E[S_1] + E[S_2|\text{not-}I] - P(I) \cdot \Delta. \quad (6)$$

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

$$E[RT_{unimodal}] = E[S_1] + E[S_2 | \text{not-}I],$$

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

$$ECI \equiv E[RT_{unimodal}] - E[RT_{crossmodal}] = P(I) \cdot \Delta. \quad (7)$$

Parameters were estimated by minimizing the Pearson χ^2 statistic

$$\chi^2 = \sum_{n=1}^4 \sum_{j=1}^2 \left(\frac{\overline{SRT}(j, n) - \widehat{SRT}(j, n)}{\sigma_{\overline{SRT}(j, n)}} \right)^2 \quad (8)$$

using the FMINSEARCH routine of MATLAB. Here $\overline{SRT}(j, n)$ and $\widehat{SRT}(j, n)$ are, respectively, the observed and the fitted values of the mean SRT to visual–auditory stimuli) presented in spatial positions (coincident, $j = 1$; disparate, $j = 2$) with SOA (referred to by $n = 1$ to 4); $\sigma_{\overline{SRT}(j, n)}$ are the respective standard errors.

For a more detailed formal presentation of the model we refer to Diederich and Colonius (2008a).

References

- Bell AH, Corneil BD, Meredith MA, Munoz DP (2001) The influence of stimulus properties on multisensory processing in the awake primate superior colliculus. *Can J Exp Psychol* 55:123–132
- 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
- Bell AH, Meredith MA, Van Opstal AJ, Munoz DP (2006) Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. *Exp Brain Res* 174(1):53–9
- Blauert J (1997) *The psychophysics of human sound localization* 2nd edn. MIT Press, Cambridge
- 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
- Corneil BD, Munoz DP (1996) The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci* 16:8193–8207
- 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 (2007a) Why two “distractors” are better than one: modeling the effect of nontarget auditory and tactile stimuli on visual saccadic reaction time. *Exp Brain Res* 179:43–54
- Diederich A, Colonius H (2007b) Modeling spatial effects in visual–tactile saccadic reaction time. *Percept Psychophys* 69(1):56–67
- Diederich A, Colonius H (2008a) Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model. *Exp Brain Res* 186:1–22
- Diederich A, Colonius H (2008b) When a high-intensity “distractor” is better than a low-intensity one: modeling the effect of an auditory or tactile nontarget stimulus on visual saccadic reaction time. *Brain Res* 1242:219–230
- Diederich A, Colonius H, Schomburg A (2008) Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia* 46:2556–2562
- Frens MA, van Opstal (1998) Visual–auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res Bull* 46:211–224
- 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
- Good MD, Gilkey RH (1996) Sound localization in noise: the effect of signal-to-noise ratio. *J Acoust Soc Am* 99:1108–1117
- Harrington LK, Peck CK (1998) Spatial disparity affects visual–auditory interactions in human sensorimotor processing. *Exp Brain Res* 122:247–252
- Heuermann H, Colonius H (2001) Spatial and temporal factors in visual–auditory interaction. In: Sommerfeld E, Kompass R, Lachmann T (eds) *Proceedings of the 17th meeting of the international society for psychophysics*. Pabst Science, Lengerich, pp 118–123
- Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994) Visual–auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform* 20: 131–153
- Hughes HC, Nelson MD, Aronchick DM (1998) Spatial characteristics of visual–auditory summation in human saccades. *Vis Res* 38:3955–3963
- Kadunce DC, Vaughan JW, Wallace MT, Benedek G, Stein BE (1997) Mechanisms of within- and cross-modality suppression in the superior colliculus. *J Neurophysiol* 78:2834–2847
- King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurons in the guinea-pig superior colliculus. *Exp Brain Res* 60:492–500
- Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cogn Brain Res* 16:468–478
- Lorenzi C, Gatehouse S, Lever C (1999) Sound localization in normal-hearing listeners. *J Acoust Soc Am* 99:1810–1820
- Lueck CJ, Crawford TJ, Savage CJ, Kennard C (1990) Auditory–visual interaction in the generation of saccades in man. *Exp Brain Res* 82:149–157
- Ma WJ, Pouget A (2008) Linking neurons to behavior in multisensory perception: a computational review. *Brain Res* 1242:4–12
- McIlwain JT (1986) Effect of eye position on saccades evoked electrically from superior colliculus of alert cats. *J Neurophysiol* 55:97–112
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 10:3215–3229
- Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56:640–662
- Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol* 75:1843–1857
- Middlebrooks JC, Knudsen EI (1984) A neural code for auditory space in the cat’s superior colliculus. *J Neurosci* 4:2621–2634
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14:247–279
- Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn Brain Res* 17:154–163
- Munoz DP, Fecteau JH (2002) Vying for dominance: dynamic interactions control visual fixation and saccadic initiation in the superior colliculus. *Prog Brain Res* 140:3–19
- Munoz DP, Schall JD (2004) Concurrent, distributed control of saccade initiation in the frontal eye field and superior colliculus.

- In: Hall WC, Moschovakis A (eds) *The superior colliculus: new approaches for studying sensorimotor integration*. CRC Press, Boca Raton, pp 55–82
- Populin LC, Yin TC (2002) Bimodal interactions in the superior colliculus of the behaving cat. *J Neurosci* 22:2826–2834
- Raab DH (1962) Statistical facilitation of simple reaction times. *Trans N Y Acad Sci* 24:574–590
- Rowland BA, Quessy S, Stanford TR, Stein BE (2007) Multisensory integration shortens physiological response latencies. *J Neurosci* 27(22):5879–5884
- Rowland BA, Stein BE (2007) Multisensory integration produces an initial response enhancement. *Front Integr Neurosci* 1:4. doi: [10.3389/neuro.07.004.2007](https://doi.org/10.3389/neuro.07.004.2007)
- Sparks DL (1999) Conceptual issues related to the role of the superior colliculus in the control of gaze. *Curr Opin Neurobiol* 9:698–707
- Sparks DL, Freedman EG, Chen LL, Gandhi NJ (2001) Cortical and subcortical contributions to coordinated eye and head movements. *Vis Res* 41:3295–3305
- Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. *Curr Biol* 13:R519–R521
- Steenken R, Colonus H, Diederich A, Rach S (2008) Visual–auditory interaction in saccadic reaction time: effects of auditory masker level. *Brain Res* 1220:150–156
- Stein BE (1998) Neural mechanisms for synthesizing sensory information and producing adaptive behavior. *Exp Brain Res* 123: 124–135
- Stein BE, Meredith MA (1993) *The merging of the senses*. The MIT Press, Cambridge
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17(4):962–974
- 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*, Cambridge, MIT Press, 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
- Wightman FL, Kistler DJ (1989) Headphone simulation of free field listening I: stimulus synthesis. *J Acoust Soc Am* 85:858–867