# A Stochastic Model of Multimodal Integration in Saccadic Responses

Hans Colonius[1] and Adele Diederich[2]

[1] Oldenburg University, P.O. Box 2503, D-26111 Oldenburg, Germany,
`hans.colonius@uni-oldenburg.de`,
`http://www.psychologie.uni-oldenburg.de/hans.colonius/index.html`
[2] International University Bremen, School of Humanities and Social Sciences,
Campus Ring 1, 28759 Bremen
`a.diederich@iu-bremen.de`,
`http://www.iu-bremen.de/hss/adiederich/`

**Abstract.** The time-window-of-integration model is a quantitative framework for describing crossmodal effects in saccadic response time. It distinguishes a first stage of parallel peripheral processing followed by a second stage of multimodal integration. The occurrence of crossmodal effects (facilitation/inhibition) hinges upon the peripheral processes terminating within a temporal window of integration. The window mechanism is determined by unimodal stimulus properties like intensity, while the size of the effect is modulated by crossmodal stimulus properties like spatial configuration.

## 1 Multimodal Integration in Saccadic Responses

Saccades are fast, voluntary movements of the eyes to align the high-resolution fovea with objects and events of interest. In a natural environment saccades are part of a rapid goal-directed orienting response system to stimuli occurring in the periphery. Stimuli are usually multimodal: in addition to visual and auditory inputs, vestibular and somatosensory afferents have access to the saccade-generating mechanism. Thus, the oculomotor system has become a prominent site for the analysis of crossmodal integration.

For example, it has been found that saccadic reaction time to visual targets (the time between the onset of the visual stimulus and the onset of the saccadic eye movement) tends to be faster when auditory stimuli are presented in close temporal or spatial proximity (see [1], [2], [3], [4]). Similar response enhancement effects for saccades have been observed for combining visual and somatosensory stimuli (cf. [5] for monkeys; [6] for humans).

These behavioral studies are in line with neurophysiological evidence for multisensory integration in the deep layers of the superior colliculus (DLSC) (see [7], [8]). Multisensory neurons in DLSC of anaesthetized cats ([9]) and monkeys ([10]) showed an enhanced response to particular combinations of visual, auditory, and tactile stimuli paralleling the spatial-temporal rules in the behavioral studies. Similar results for recordings from the awake behaving monkey have recently been obtained ([11], [12]).

Here we present a stochastic model that establishes a formal framework in which rules of crossmodal integration can be stated. Within that framework one can specify how the integration mechanism depends on the uni- and multimodal stimulus parameters and on specifics of the experimental paradigm. It allows to make qualitative and quantitative predictions and should thereby ultimately provide a link between the neural and the behavioral level of investigation.

## 2 The Time-Window-of-Integration (TWIN) Model

Since stimulation from different modalities like vision and touch cannot interact (e.g., on the retina), the model claims the existence of a *first stage* of parallel independent modality-specific activations in the afferent pathways. 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. This does not preclude the possibility of interaction between modality-specific pathways, nor between modality-specific and crossmodal areas, at a later stage of processing. 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 ([13]). Thus the entire processing time must consist of at least two stages arranged in series. The *second stage* comprises neural integration of the input and preparation of the ocular motor response. True interaction, however, resulting in facilitation or inhibition of the response is supposed to occur only if the peripheral processes of the first stage all terminate within a given temporal "window of integration".

Even under invariant experimental conditions, saccadic responses typically vary from one trial to the next due to an inherent variability of the underlying neural processes in both ascending and descending pathways. This is taken into account by assuming the duration of each of the stages to be a random variable.

### 2.1 Distribution-Free Model Properties and Predictions

According to the model, observed reaction time in the multimodal condition can be written as a sum of two nonnegative random variables with finite first and second moments:

$$RT_{multimodal} \overset{d}{=} W_1 + W_2, \tag{1}$$

where $W_1$ and $W_2$ refer to first and second stage processing time, respectively[3]. Let $I$ denote the event that crossmodal interaction occurs, having probability $\mathrm{P}[I]$. Thus the saccadic response time (SRT) distribution is a binary mixture of two distributions defined by conditioning on event $I$:

---

[3] $\overset{d}{=}$ stands for "equal-in-distribution".

$$\begin{aligned}\mathrm{P}[RT_{multimodal} \le t] &= \mathrm{P}[W_1 + W_2 \le t] \\ &= \mathrm{P}[I]\mathrm{P}[W_1 + W_2 \le t|I] + (1 - \mathrm{P}[I])\mathrm{P}[W_1 + W_2 \le t|\text{not-}I].\end{aligned} \tag{2}$$

While neither $\mathrm{P}[I]$ nor the two conditional distributions in Eq.(2) can be estimated directly from the data, mixture distributions have several distinctive properties that lead to empirically testable predictions even if no specific distribution assumptions are introduced (see [14]).

For the expected SRT in the multimodal condition then follows:

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

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

$$\mathrm{E}[RT_{multimodal}] = \mathrm{E}[W_1] + \mathrm{E}[W_2|\ \text{not-}I] - \mathrm{P}[I]\ \Delta. \tag{3}$$

The product $\mathrm{P}[I]\ \Delta$ is a measure of the expected crossmodal interaction in saccadic RT in the second stage, with positive $\Delta$ values corresponding to facilitation, negative ones to inhibition.

In the unimodal conditions, no interaction is possible. Thus,

$$\mathrm{E}[RT_{unimodal}] = \mathrm{E}[W_1] + \mathrm{E}[W_2|\text{not-}I], \tag{4}$$

and the amount of crossmodal interaction is

$$\mathrm{E}[RT_{unimodal}] - \mathrm{E}[RT_{multimodal}] = \mathrm{P}[I]\ \Delta.$$

Several empirically testable predictions can now be formulated. First, the amount of crossmodal interaction should depend on the stimulus onset asynchrony (SOA) between the stimuli. For example, 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. Indeed, the effect of crossmodal interaction tends to be most prominent when there is some characteristic temporal asynchrony between the stimuli ([1],[2]). Second, the probability of interaction, $\mathrm{P}[I]$, 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 $\mathrm{P}[I]$ is in line with the empirical observation that a very strong target signal will effectively suppress any interaction with other modalities. The principle of "inverse effectiveness", according to which crossmodal facilitation is strongest when stimulus strengths are weak or close to threshold level ([9]), 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 crossmodal integration. Third, the amount of crossmodal interaction ($\varDelta$) 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[4]. On the other hand, crossmodal features have no influence on first stage processing time since the modalities are yet being processed in separate pathways.

More specific predictions concerning the effects of varying stimulus intensity are implied by the following rules governing the window-of-integration mechanism. When the task is to orient toward the target modality stimulus ignoring stimuli from other modalities (*focused attention*), the first stage duration is determined by the target peripheral process, but crossmodal integration is occurring only if the non-target stimulus wins the race in the first stage, i.e., the window of integration is opened only by activity triggered by the non-target stimulus. Increasing the intensity of the target stimulus will thus increase its chances to win the race decreasing the probability that the window of integration opens, so that less crossmodal interaction should occur. This prediction is in line with the observation that a very strong target signal will suppress any interaction with other modalities. Increasing the intensity of the non-target stimulus, however, leads to the opposite prediction: the non-target stimulus will have a better chance to win the race and to open the window of integration, hence predicting more crossmodal interaction on average. On the other hand, when the task is to orient toward the first stimulus detected no matter of which modality (*redundant target*), 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. Here, the effect of stimulus intensity depends on additional assumptions not outlined here.

### 2.2 TWIN Model with Exponential First Stage Distributions

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. [15]). To illustrate the derivation for the expected SRT, consider a focused attention experiment with a visual target and an auditory non-target stimulus. The first stage duration is determined by the target peripheral process of random duration $V$, say, yielding $\mathrm{E}[W_1] = \mathrm{E}[V] = 1/\lambda_V$ ($\lambda_V$ denotes the intensity parameter of the exponential distribution of $V$). From the assumptions stated in the last section,

$$I = \{A + \tau < V < A + \tau + \omega\} \tag{5}$$

---

[4] Laterality here means whether or not all stimuli appear in the same hemisphere.

where $A$ is the peripheral auditory latency and $\tau$ and $\omega$ denote SOA and window width, resp. Straightforward calculation yields

$$\mathrm{P}[I] = \frac{\lambda_A}{\lambda_A + \lambda_V} \{\exp[-\lambda_V \tau] - \exp[-\lambda_V(\tau + \omega)]\}\,, \tag{6}$$

where $\lambda_A$ refers to the auditory intensity parameter. It is obvious from Eq. (6) that the probability of interaction increases both with $\lambda_A$ and the window width $\omega$, as it should. Expected saccadic reaction time then is

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

where $\mu = \mathrm{E}[W_2|\text{not-}I]$, 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 $\varDelta$ is positive: in any given trial, whenever the visual peripheral process ($V \equiv W_1$) is relatively slow, the auditory 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.

## 3 Conclusion

The TWIN model has recently been shown to give an excellent description of crossmodal effects on SRT in visual-auditory and visual-tactile focused attention experiments ([1],[6]). Note, however, that it is not meant to mirror multisensory processes at the level of an individual neuron. There are many different types of multisensory convergence occurring in individual neurons (see [16]), and some of their activities are consistent with the TWIN assumptions while others are not. Note also 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 future work, the second stage mechanisms should be specified in more detail, 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. [17]), and connecting these with behavioral data via an appropriate elaboration of the TWIN model should be a challenging task.

*Notes and Comments.* This work was supported by DFG grants Di 506/7-2 and SFB 517/C3 (Neurokognition).

## References

1. Colonius, H., Arndt, P.: A two-stage model for visual-auditory interaction in saccadic latencies. Perc. Psychophys. **63** (2001) 126–147
2. Frens, M.A., Van Opstal, A.J., Van der Willigen, R.F.: Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. Perc. Psychophys. **57** (1995) 802–816
3. Harrington, L.K., Peck, C.K.: Spatial disparity affects visual-auditory interactions in human sensorimotor processing. Exp. Brain Res. **122** (1998) 247–252
4. Hughes, H.C., Nelson, M.D., Aronchick, D.M.: Spatial characteristic of visual-auditory summation in human saccades. Vis. Res. **38** (1998) 3955–3963
5. Groh, J.M., Sparks, D.L.: Saccades to somotosensory targets. I. Behavioral characteristics. J. Neurophys. **75** (1996) 412–427
6. Diederich, A., Colonius, H., Bockhorst, D., Tabeling, S.: Visual-tactile interaction in saccade generation. Exp. Brain Res., to appear
7. Munoz, D.P., Wurtz, R.H.: Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. J. Neurophysiol. **73** (1995a) 2313–2333
8. Munoz, D.P., Wurtz, R.H.: Saccade-related activity in monkey superior colliculus. II. Spread of activity during saccades. J. Neurophysiol. **73** (1995b) 2334–2348
9. Meredith, M.A., Stein, B.E.: Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol. **56** (1986) 640–662
10. Wallace, M.T., Wilkinson, L.K., Stein, B.E.: Representation and integration of multiple sensory inputs in primate superior colliculus. J. Neurophysiol. **76** (1996) 1246–1266
11. Bell, A.H., Corneil, B.D., Meredith, M.A., Munoz, D.P.: The influence of stimulus properties on multisensory processing in the awake primate superior colliculus. Can. J. Exp. Psych. **55:2** (2002) 123–132
12. Frens, M.A., Van Opstal, A.J.: Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. Brain Res. Bull. **46** (1998) 211–224
13. Driver, J., Spence, C.: Multisensory perception: Beyond modularity and convergence. Curr. Biol. **10** (2000) 731–735
14. Yantis, S., Meyer, D.E., Smith, J.E.K.: Analysis of multinomial mixture distributions: New tests for stochastic models of cognition and action. Psych. Bull. **110** (1991) 350–374
15. Luce, R.D.: Response Times: Their Role in Inferring Elementary Mental Organization. Oxford University Press, New York (1986)
16. Meredith, M.A.: On the neural basis for multisensory convergence: a brief overview. Cog. Brain Res. **14** (2002) 31–40
17. Kadunce, D.C., Vaughan, J.W., Wallace, M.T., Stein, B.E. The influence of visual and auditory receptive field organization on multisensory integration in the superior colliculus. Exp. Brain Res. **139** (2001) 303–310