

The time window of multisensory integration: relating reaction times and judgments of
temporal order

Adele Diederich

Jacobs University Bremen and

Research Center Neurosensory Science, Oldenburg University, Germany

Hans Colonius

Department of Psychology and Cluster of Excellence “Hearing4all”,

Oldenburg University, Germany

Abstract

Even though visual and auditory information of one and the same event often do not arrive at the sensory receptors at the same time, due to different physical transmission times of the modalities, the brain maintains a unitary perception of the event, at least within a certain range of sensory arrival time differences. The properties of this “temporal window of integration”, its recalibration due to task requirements, attention, and other variables, have recently been investigated intensively. Up to now, however, there has been no consistent definition of “temporal window” across different paradigms for measuring its width. Here we propose such a definition based on our time-window-of-integration (TWIN) model [Colonus & Diederich, *J Cog Neurosci* 16, 1000–10, 2004]. It applies to judgments of temporal order (or simultaneity) as well as to reaction time paradigms. Reanalyzing data from Mégevand et al. [*PLoS ONE*, 8 (8), e71608, 2013] by fitting the TWIN model to data from both paradigms, we confirm the authors’ hypothesis that the temporal window in a reaction time task tends to be wider than in a temporal order judgment task. This first step toward a unified concept of “temporal window of integration” should be a valuable tool in guiding investigations of the neural and cognitive bases of this so far somewhat elusive concept.

Key words: multisensory integration, temporal window of integration, temporal order judgment, reaction time, recalibration

The time window of multisensory integration: relating reaction times and judgments of temporal order

Human adaptive behavior depends on the ability of the perceptual system to rapidly deliver information about on-going events in the environment. This information typically arrives in parallel via different sensory channels; in order to achieve a coherent and valid perception of the outside world the brain must determine which of these temporally coincident sensory signals are caused by the same physical source and should thus be integrated into a single percept (Koerding et al., 2007). This task is made more difficult by the fact that there are subtle differences in arrival times, for example, of sound and light: For a synchronized audiovisual event occurring in the near field, audition will be perceived before vision by about 30 ms because neural transduction for audition is much faster than for vision (Alais & Carlile, 2005). This more than compensates for the slower physical speed of sound. For events occurring beyond a distance¹ of 10 to 15 m, a visual stimulus will be perceived first and increasingly so as distance increases, due to the much faster speed of light. Nevertheless, in daily life we are typically not aware of these subtle differences in arrival times of sound and light and perceive the stimuli as simultaneous. The range of arrival time differences the brain tolerates in treating the two information streams as belonging to the same event has sometimes been termed the *temporal (binding) window of multisensory integration* (Dixon & Spitz, 1980; Spence & Squire, 2003; Colonius & Diederich, 2004; Vroomen & Keetels, 2010) . The exact size of this window, its potential malleability and dependence on stimulus modalities, task demands, and individual differences has recently been in the focus of numerous studies in multisensory research (e.g., Stevenson et al., 2014; Stevenson, Zemtsov, & Wallace, 2012; Mégevand, Molholm, Nayak, & Foxe, 2013; Van Wassenhove, Grant, & Poeppel, 2007; van Eijk, Kohlrausch, Juola, & van de Par, 2009; Powers, Hillock, & Wallace, 2009; Russo et al., 2010; Hillock, Powers, & Wallace, 2011; Hillock-Dunn & Wallace, 2012; de Boer-Schellekens & Vroomen,

¹sometimes called ‘horizon of simultaneity’ (Pöppel, Schill, & von Steinbüchel, 1990)

2014; Magnotti, Ma, & Beauchamp, 2013; van Wanrooij, Bell, Munoz, & van Opstal, 2009; Corneil, van Wanrooij, Munoz, & van Opstal, 2002).

The concept of a temporal binding window of multisensory integration has been discussed primarily in two experimental paradigms presenting stimuli from different modalities. The first one, the redundant signals reaction time (RT) task, compares unimodal and multimodal reaction time (RT) in order to derive an index of multisensory integration. The second asks participants to report which stimulus they perceived first (temporal order judgment, TOJ); sometimes, judgment of simultaneity (SJ) is elicited, either in addition or instead of TOJ. In a recent audiovisual study, Mégevand and colleagues (Mégevand et al., 2013) probed whether the size of the temporal window was invariant across audiovisual RT and TOJ tasks using identical stimuli in both paradigms. Temporal windows in the RT task turned out to be wider than in the corresponding TOJ task. This was consistent with the authors' hypothesis, namely, that in the latter task where participants have to discern small asynchronies between the acoustic and the visual stimulus, the temporal window is set to a value as narrow as possible for optimal performance. On the other hand, optimal performance in the RT task "...would entail widening the window to maximize multisensory facilitation" (Mégevand et al., 2013, p. 2).

However, given that the authors' operational definition of temporal window of integration in this study differed for the two tasks (see below), there is a clear possibility that observing windows with different width in the two tasks was mainly due to this difference in definitions. The issue addressed here is how to develop a theory-based definition of temporal window that would be valid across different experimental paradigms. We suggest a first step towards developing a definition of temporal binding window comprising both reaction time and temporal order judgment paradigms.

First, within a quantitative modeling framework for multisensory reaction time (time-window-of-integration, or TWIN model) (Colonus & Diederich, 2004; Diederich & Colonus, 2004), the width of the window is represented as a numerical parameter, defined

as a certain difference in peripheral arrival times, modulating the probability of multisensory integration in the RT task. Although the probability of integration is not directly observable in the RT task, it can nevertheless be estimated by the TWIN model. Second, it will be shown how a standard auxiliary assumption relating the detection and discrimination mechanisms in TOJ and RT allows us to predict temporal order frequencies from the bimodal reaction times and thereby make inferences about the malleability of the temporal window. The next section introduces the basic features of the TWIN model, followed by an exposition of its relation to the TOJ task and the test of window malleability. Our approach is illustrated by a reanalysis of data from a study by Mégevand and colleagues. Finally, some alternative modeling approaches for TOJ and RT tasks will be discussed.

Time-window-of-integration (TWIN) model

The TWIN model (Colonius & Diederich, 2004; Diederich & Colonius, 2004) is a quantitative framework that was developed to predict the effect of the spatiotemporal parameters of a crossmodal experiment on response speed. It postulates that a crossmodal (audiovisual) stimulus triggers a race mechanism in the very early, peripheral sensory pathways (*first stage*), followed by a compound stage of converging sub-processes comprising neural integration of the input and preparation of a response. This *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 central assumption of the model concerns the temporal configuration needed for crossmodal interaction to occur:

TWIN assumption: Crossmodal interaction occurs only if the peripheral processes of the first stage all terminate within a given temporal interval, the ‘time window of integration’.

Thus, the window acts as a filter determining whether afferent information delivered from different sensory organs is registered close enough in time to trigger multisensory

integration. Passing the filter is necessary but not sufficient for crossmodal interaction to occur because the amount of interaction may also depend on many other aspects of the stimulus set, in particular the spatial configuration of the stimuli. The amount of crossmodal interaction manifests itself in an increase, or decrease, of second stage processing time. Although this amount does not directly depend on the stimulus onset asynchrony (SOA) of the stimuli, temporal tuning of the interaction occurs because the *probability of integration* is modulated by the SOA value (see next).

Deriving the probability of integration in TWIN

The race in the first stage of the model is made explicit by assigning statistically independent, nonnegative random variables V and A , say, to the processing times for a visual and an acoustic stimulus, respectively. With τ as SOA value and ω as integration window width parameter, the TWIN assumption implies that the event that multisensory integration occurs, denoted by I , equals

$$I = \{|V - (A + \tau)| < \omega\} = \{A + \tau < V < A + \tau + \omega\} \cup \{V < A + \tau < V + \omega\},$$

where the presentation of the visual stimulus is arbitrarily defined as the zero time point. 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.

Consistent with previous RT studies probing the TWIN model (Diederich & Colonius, 2008a, 2008b, 2007a, 2007b), we assume that the peripheral processing times for the visual (V) and for the acoustic (A) stimulus follow an exponential distribution with parameters λ_V and λ_A ($\lambda_V, \lambda_A > 0$), respectively. Thus, the distribution functions are

$$F_V(t) = 1 - \exp[-\lambda_V t] \quad \text{and} \quad F_A(t) = 1 - \exp[-\lambda_A t],$$

for $t \geq 0$, and $F_V(t) = F_A(t) = 0$ for $t < 0$. The computation of $P(I)$ requires evaluation² of

$$\Pr(A + \tau < V < A + \tau + \omega) = \int_0^\infty \{F_V(a + \tau + \omega) - F_V(a + \tau)\} dF_A(a)$$

²Here and below we use the more compact notation of the *Lebesgue-Stieltjes* integral.

and

$$\Pr(V < A + \tau < V + \omega) = \int_0^\infty \{F_A(v + \omega - \tau) - F_A(v - \tau)\} dF_V(v).$$

The values of these integrals (listed in the appendix) depend on the sign of τ and $\tau + \omega$ and adding them yields an explicit expression for $P(I)$.

Deriving mean RT in TWIN

Writing S_1 and S_2 for first and second stage processing time, respectively, overall expected reaction time in the crossmodal condition with an SOA equal to τ , $E[RT_{VA,\tau}]$, is computed conditioning on event I occurring or not,

$$\begin{aligned} E[RT_{VA,\tau}] &= E[S_1] + P(I) E[S_2|I] + [1 - P(I)] E[S_2|I^c] \\ &= E[S_1] + E[S_2|I^c] - P(I) \times \Delta. \\ &= E[\min(V, A + \tau)] + \mu - P(I) \times \Delta. \end{aligned} \tag{1}$$

Here, I^c denotes the complementary event to I , μ is short for $E[S_2|I]$, and Δ stands for $E[S_2|I^c] - E[S_2|I]$. The term $P(I) \times \Delta$ is a measure of the expected amount of crossmodal interaction in the second stage, with positive Δ values corresponding to facilitation, negative ones to inhibition. An explicit expression for $E[RT_{VA,\tau}]$ as a function of the parameters is found in the appendix.

Deriving the amount of crossmodal interaction in TWIN

Although the unimodal RTs are not needed in predicting the TOJ judgment probabilities, we mention them here for the sake of completeness and because they allow deriving the predicted amount of crossmodal interaction. Event I cannot occur in the unimodal (visual or auditory) condition, thus expected reaction time for these conditions is, respectively,

$$E[RT_V] = E[V] + E[S_2|I^c] \quad \text{and} \quad E[RT_A] = E[A] + E[S_2|I^c].$$

Note that the race in first stage produces a (not directly observable) statistical facilitation effect (*SFE*) analogous to the one in the “classic” race model (Raab, 1962):

$$SFE \equiv \min\{E[V], E[A] + \tau\} - E[\min\{V, A + \tau\}].$$

This contributes to the overall crossmodal interaction effect predicted by TWIN which amounts to:

$$\min\{E[RT_V], E[RT_A] + \tau\} - E[RT_{V,A,\tau}] = SFE + P(I) \times \Delta.$$

Thus, crossmodal facilitation observed in a redundant signals task may be due to either multisensory integration or statistical facilitation, or to both. Moreover, a potential multisensory inhibitory effect occurring in the second stage may be weakened, or even masked completely, by simultaneous presence of statistical facilitation in the first stage.

From RT to TOJ: An auxiliary assumption

There is a long history of studies on the relation between reaction times and temporal order judgments (Rutschmann & Link, 1964; Neumann, Esselmann, & Klotz, 1993; Neumann & Niepel, 2004; Jáskowski, Jaroszyk, & Hojan-Jeziarska, 1990); for a recent review see Miller and Schwarz (2006). An influential model of how these measures relate is based on the idea that both RT and TOJ depend on the duration of an initial, perceptual detection stage that is identical for both tasks (Gibbon & Rutschmann, 1969). This model is referred to by Miller and Schwarz (2006) as the *canonical model* (ibid, p. 394). As the TWIN model was originally not developed to account for data from TOJ, we will adopt the canonical model as an auxiliary assumption here. In particular, in this *extended* TWIN model we assume (i) that subjects’ temporal order judgments are based on the first-stage processing times, V and A , representing the time to detect the visual and the acoustic stimulus in the RT task, respectively, and (ii) that these judgments are modulated by the time window of integration in a way to be specified below.

Typically, data from an audiovisual TOJ task are presented in the format of the (relative) frequency of responding “visual stimulus first” as a function of the stimulus onset asynchrony τ between visual and acoustic stimulus, yielding an estimate for the psychometric function

$$\Psi(\tau) = \Pr(\text{“visual first”} \mid \text{visual stimulus presented } \tau \text{ ms before the acoustic}).$$

Thus, for $\tau < 0$ the acoustic stimulus is presented τ ms before the visual and for $\tau > 0$ the order is reversed. Under the canonical model, with visual and acoustic stimuli that are physically identical in both tasks, we assume that whenever the detection times for the stimuli fall within the integration window, subjects cannot base their judgment on sensory evidence about the arrival times and hence respond “visual first” with fixed (bias) probability, β , say. If the stimulus arrival times do not fall within the time window of integration, then the response is based on the temporal order of detection proper. Considering the relevant arrival time events, we get

$$\begin{aligned} \Psi(\tau) &= \Pr(I) \times \beta + \Pr(V + \omega < A + \tau) \\ &= \{\Pr(A + \tau < V < A + \tau + \omega) + \Pr(V < A + \tau < V + \omega)\} \times \beta \\ &\quad + \Pr(V + \omega < A + \tau). \end{aligned} \tag{2}$$

If no a-priori response bias exists, β will equal 0.5. For the computation of the detection order probabilities under the TWIN model, the cases of $\tau < 0$ and $\tau > 0$ must be considered separately, denoted here as $\Psi^-(\tau)$ and $\Psi^+(\tau)$, so that for any real-valued τ

$$\Psi(\tau) = \begin{cases} \Psi^+(\tau), & \text{if } \tau \geq 0; \\ \Psi^-(\tau) & \text{if } \tau < 0. \end{cases}$$

For the exponential version of TWIN, the psychometric function and its derivation are found in the appendix.

Figure 1 presents the probability of a “visual first” response, i.e., psychometric function Ψ , as a function of SOA (τ) and the response bias (β), with window size $\omega = 200$

[ms] and $\lambda_V = 1/100$, $\lambda_A = 1/50$. Unsurprisingly, Ψ is increasing in both arguments, τ and β . Moreover, the SOA range of steepest ascent is modulated by β : for small β , the probability of a “visual first” response depends mainly on the visual winning against the auditory with a lead of ω ms, i.e., without the time window opening, which only happens when the auditory is delayed long enough (i.e., large τ). For β increasing towards one, this effect weakens allowing the probability to ascend steeply for smaller SOAs.

[Figure 1 about here]

Testing the malleability of the temporal window of integration

In the canonical model the stimulus parameters λ_A and λ_V are assumed to be identical for both tasks, RT and TOJ. Probing malleability of the temporal window thus amounts to asking whether or not this invariance also holds for the window width parameters across the two tasks. A straightforward way of probing this is to fit the extended TWIN model to the RT and TOJ data simultaneously and to investigate whether the two parameters, ω_{RT} and ω_{TOJ} , differ significantly. This would involve comparing χ^2 goodness-of-fit criteria with one versus two estimates for the window width.

Although a recent parameter recovery study (Kandil, Diederich, & Colonius, 2014) has revealed that the parameters of the TWIN model, including window width, can be recovered with high accuracy and precision, a test not being based on a single set of parameter estimates may be preferable here. Thus, we suggest an alternative approach based on a nonparametric bootstrap method (e.g., Davison & Hinkley, 1997): Given the set of empirical RT and TOJ data, the extended TWIN model is repeatedly fit to N samples, drawn with replacement from that data set, generating N pairs of parameter estimates $(\omega_{RT}, \omega_{TOJ})$. The distribution of these values gives a measure of the variability and direction of the difference $(\omega_{RT} - \omega_{TOJ})$. The decision on whether the two window estimates are equal or one is smaller than the other, is then based on the nonparametric Wilcoxon signed-rank test (Hollander & Wolfe, 1999) applied to these differences. We

illustrate this approach on the set of data from Mégevand et al. (2013) with $N = 1,000$.

The Mégevand et al. (2013) study. A visual stimulus (red-colored disk) and an acoustic stimulus (1 kHz sine wave) were presented for 10 ms with SOAs of 0, ± 20 , ± 40 , ± 60 , ± 80 , ± 100 , ± 120 , ± 150 , ± 200 , ± 250 , ± 300 , ± 400 ms (negative SOAs indicating that the acoustic stimulus preceded the visual stimulus). Both stimuli were also presented unimodally. Data from 11 participants were retained for analysis³. Data from the RT task were tested for violations of the race model inequality (RMI) (Raab, 1962; Miller, 1982, 1986; Colonius & Diederich, 2006). This “classic” model assumes that the response in the bimodal condition is determined by the “winner of a race” between the sensory-specific channels (“separate activation” assumption). For each SOA the inequality compares the distribution function of RTs in the bimodal condition with the sum of the two unimodal distribution functions. If certain assumptions are met (Colonius, 1990), a violation of the inequality indicates that the speed-up of RTs in the bimodal condition is greater than predicted by a simple probability summation model (“statistical facilitation”, *SFE*). Using a conservative test method – a resampling procedure suggested in Gondan (2010) – all but one participant showed a significant violation of the inequality at $\text{SOA} = 0$ ($p < .05$), and none of them displayed a violation beyond the ± 120 range. Violations were more common with the visual stimulus leading the acoustic one. Separately for each participant, the temporal window of integration based on RT was defined by the contiguous SOAs with significant violations of the RMI that were around physical simultaneity or closest to it (Mégevand et al., 2013, p. 3). This definition is based on the notion that violations of the RMI at a specific SOA value are due to both unisensory processes falling into a temporal window so that multisensory integration speeds up bimodal responses beyond what can be achieved by statistical facilitation alone. For the TOJ task, audiovisual stimulus pairs were presented with the same set of SOAs as for the RT task. Logistic psychometric functions were fitted to participants’ proportion of “visual first” responses across SOA in a Bayesian

³For further details of the experiment and data analysis we refer to Mégevand et al. (2013)

analysis (using a Markov chain Monte Carlo algorithm for estimation of the posterior distribution of the parameters). Above-chance performance in the TOJ task was defined by the upper and lower time points on the SOA axis where performance was at the 75% correct level (with a correction in case of lapses) yielding a temporal window defined as the corresponding range on the proportion of “visual first” responses. This definition of temporal window is based on the notion that highly accurate discrimination of visual and auditory arrival times is only possible outside of that window. Consistent with the authors’ hypothesis, there were 5 (out of 11) participants showing above-chance TOJ performance at SOA values where significant RMI violations were observed, and these participants had narrower TOJ-defined windows compared to the other subjects, whereas the widths for the RT-defined windows did not differ between these two groups.

TWIN reanalysis of Mégevand et al. (2013)

The data set consists of 46 experimental conditions, defined by the 23 SOA values for both the RT and TOJ tasks and ignoring the unimodal RT data. About 60 observations were collected under each condition. Separately for each participant, the extended TWIN model was fitted by minimizing⁴ the objective function.

$$\text{Objective Function} = \sum_{\tau} \left\{ \left(\frac{\text{mean}[RT_{VA,\tau}] - E[\widehat{RT}_{VA,\tau}]}{\text{standard error}(\text{mean}[RT_{VA,\tau}])} \right)^2 + \left(\frac{f_{VA,\tau} - \widehat{\Psi}(\tau)}{\text{standard error}[f_{VA,\tau}]} \right)^2 \right\},$$

where (i) $\text{mean}[RT_{VA,\tau}]$ and $f_{VA,\tau}$ refer to mean RTs and relative frequency of “visual first” judgments in the RT and TOJ task, respectively, (ii) the expressions with a hat are the predicted values, and (iii) summation is over all SOA values (τ). This function measures the deviation of the RT means and TOJ relative frequencies from the corresponding model predictions. Minimization was performed with respect to parameter estimates of λ_V , λ_A , μ , Δ , β , ω_{RT} , and ω_{TOJ} . Table 1 contains the estimates for all 11 participants. Figure 2 shows an example fit of the extended TWIN model for mean bimodal RTs and the TOJ psychometric function across the SOA range.

⁴Function `fminsearch` in MATLAB[®]. The program is freely available from the authors upon request.

[Figure 2 about here]

Subsequently, the extended TWIN model was fitted to each of $N = 1,000$ bootstrap samples from the joint RT and TOJ data using the same objective function as before. Table 1 (last column) contains the estimates for all 11 participants of the 99%-confidence intervals for the (pseudo)median⁵ of the distribution of the bootstrapped time window differences, $\omega_{RT} - \omega_{TOJ}$. The results are quite clear-cut: for all but two participants (5 and 8), the time window for the RT task is larger than for the TOJ task ($p < .001$, Wilcoxon signed-rank test).

[Table 1 about here]

Thus, this result supports the hypothesis of Mégevand and colleagues. Moreover, it goes beyond their findings because all 11 participants, instead of only 5, have been revealed to show malleability of the temporal window, with nine of them consistent with the hypothesis and the other two showing larger temporal windows for the TOJ than for the RT condition. The magnitudes of the window widths found here, however, are quite different from those in Mégevand et al. (2013): only for three participants the window width parameters were in the ± 120 ms SOA range where violations of the race model inequality had been found by Mégevand et al. (2013). This is not surprising given the difference between the notion of temporal window within the TWIN model and an SOA range-defined temporal window (see below).

Because our main interest here is the comparison of the time window widths in RT and TOJ tasks, we refrain from discussing in detail the goodness of fit of the extended TWIN model to the Mégevand et al. (2013) data. For five of the subjects the model could be rejected under a χ^2 criterion ($p < .05$) but all qualitative features of both RT and TOJ data were well described for all 11 subjects.

⁵This location parameter is the median of the midpoints of pairs of observations estimated by the Hodges-Lehmann statistic (Hollander & Wolfe, 1999); it is equal to the median for symmetric distributions.

Discussion

The temporal window of integration has become an important conceptual tool in describing crossmodal binding effects in a variety of multisensory integration tasks (for a recent review, see Chen & Vroomen, 2013). Its exact definition, however, has remained specific to the specific task being studied, and this lack of generality makes it difficult to compare results collected from judgments of temporal order, for example, with those from measuring crossmodal reaction times when it comes to testing hypotheses about the width of the window. The main result here is to have introduced a common theoretical basis for the temporal window concept across these two rather different paradigms by tying it to a common model.

Specifically, we have suggested an extension of the time window of integration (TWIN) model of Colonius and Diederich (2004) to simultaneously assess the size of the window for reaction time and temporal order judgment, given the same set of stimuli. In this extension, window width emerges as a model parameter controlling, on the one hand, the probability of crossmodal interaction occurring in reaction time and, on the other, the probability of judging the temporal order of the stimuli. In the TOJ task, the width of the window determines how often the two stimuli will be “bound together” and, thereby, how often the subject can only guess that the visual stimulus occurred first, requiring the introduction of a response bias parameter β into the model. The viability of this proposal is illustrated by a reanalysis of data from Mégevand et al. (2013), supporting and extending their hypothesis of a smaller time window for the TOJ task compared to the crossmodal RT task.

Although our findings are consistent with the Mégevand et al. (2013) hypothesis, there is a notable difference between the time window concepts: time window width in TWIN is a numerical parameter that determines how close the random arrival times between peripheral visual and auditory processes must be in order to trigger multisensory integration, thereby preventing a stimulus-based order judgment. This is in stark contrast

to the definition of time window in Mégevand et al. (2013) and many other studies, where the notion is always tied to a specific physical SOA point or range (for a similar point distinguishing objective and subjective SOAs, see Yarrow, Jahn, Durant, & Arnold, 2011). Because of this fundamental difference there is no point in numerically comparing the time window widths obtained in the two approaches. Nevertheless, the observed consistency of our results with those of Mégevand et al. (2013) suggests that both do capture some common aspects of the underlying processes.

Results obtained with the TOJ task typically differ from those using simultaneity judgments with two response alternatives (SJ2) or with three response alternatives (SJ3) (e.g., van Eijk, Kohlrausch, Juola, & van de Par, 2008; Ulrich, 1987; Sternberg & Knoll, 1973). It would be straightforward to extend our approach to these simultaneity judgment tasks. In fact, in an investigation to explain the empirical differences among these tasks, García-Pérez and Alcalá-Quintana (2012) developed a model to disentangle the sensory and decisional components in the three different tasks in such a way that special cases of their model would reduce to the judgment part of the extended TWIN model introduced here.

Recalibration: fast, slow, and asymmetric

Widening the temporal window of integration in a RT task, or narrowing it in a TOJ task, can be seen as an observer's strategy to optimize performance in an environment where the temporal structure of sensory information from separate modalities provides a critical cue for inferring the occurrence of crossmodal events (for a recent review, see Vroomen & Keetels, 2010). Often, this temporal recalibration is seen as perceptual learning resulting from some, or even extended, training in TOJ or SJ tasks (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Powers et al., 2009; Powers, Hevey, & Wallace, 2012; Navarra, Hartcher-O'Brien, & Spence, 2009; Harrar & Harris, 2008) and, recently, effects of week-long synchronous and asynchronous adaptation conditions on reaction times to audiovisual stimuli have been found (Harrar, Spence, & Harris, 2013).

On the other hand, rapid recalibration taking place from one trial to the next, would clearly be advantageous in a dynamically changing environment. This has actually been observed within an *auditory localization* task, where spatial recalibration occurred as a function of audiovisual discrepancy after a single trial presentation (Wozny & Shams, 2011; Mendonca, Escher, van de Par, & Colonius, 2014). Interestingly, *temporal* recalibration has also been found in a recent audiovisual study with randomly changing SOA values (van der Burg, Alais, & Cass, 2013). Their participants experienced luminance onsets presented 35 ms before the tone’s onset as “synchronous” when vision occurred first on the previous trial and independent of whether the previous trial was perceived as “synchronous” .

Van der Burg and colleagues also registered a strong asymmetry: the shift of the point of subjective simultaneity was much smaller, and only within a shorter SOA range, when audition occurred first on the previous trial. A recent magnetoencephalography study (Kösem & Van Wassenhove, 2013), where participants’ perceived simultaneity could be accounted for by systematic shifts in the phase of auditory but not visual neural responses, suggests that auditory cortex recalibrates its timing to the visual spatial anchor.

While further research on the neural mechanisms underlying both fast and slow recalibration and its asymmetry is called for, the TWIN modeling framework also needs to be extended in order to capture dependencies across trials. Trial-to-trial recalibration could be represented, for example, by assuming a different value for the time window parameter, depending on whether the previous trial had an acoustic stimulus leading the visual, or vice versa. This across-trial asymmetry is to be distinguished from within-trial asymmetry occurring, e.g., in psychometric functions of simultaneity judgement (Powers et al., 2009). Notably, no modification of TWIN is required to account for this latter type of asymmetry.

Alternative models

Alternative approaches to simultaneously accommodate RT and TOJ data have focused on explaining the dissociation between the two measures often found in empirical

data. This dissociation occurs when, for example, increases of stimulus intensity produce reductions in RT that are not equally reflected in the psychometric functions of the TOJ task (Sanford, 1971; Jáskowski, 1992). One attempt to understand this dissociation, still within the canonical model, is the *criterion-shift hypothesis* assuming that subjects use a higher criterion for stimulus detection in RT tasks than in TOJ tasks (Sanford, 1974).

Together with the assumption that evidence for the occurrence of a stimulus accumulates faster for a more intense stimulus than for a weaker one (e.g. Grice, 1968), it can be shown that this predicts a larger effect on stimulus detection time for RT than for TOJ (see Miller & Schwarz, 2006, p. 397). Models allowing for detailed quantitative predictions of RT based on the concept of *stochastic diffusion processes* abound (Laming, 1968; Ratcliff, 1978; Diederich, 1992; Busemeyer & Townsend, 1993; Diederich, 1994, 1995; Schwarz, 1994; Smith, 1990) but had, until recently, not been developed for TOJ tasks. However, Miller and Schwarz (2006) implemented the criterion-shift hypothesis in a diffusion model approach for RT and TOJ, with no specific mechanism predicting crossmodal interaction effects. One of the authors (Schwarz, 2006) developed an alternative RT/TOJ diffusion model specifically addressing the *redundant signals effect*, i.e., the facilitation of responses to redundant stimuli – either from the same or from different modalities – compared with the responses to single stimuli. The RT part of this model consists of a superposition of diffusion processes for redundant stimuli (Schwarz, 1994) (see also Diederich, 1992, 1994) while, for the TOJ part, a differencing rule is postulated, i.e., the observer is assumed to monitor the ongoing difference in sensory activation induced by the two stimuli over time and the TOJ psychometric function is determined by the probability of the difference first crossing the upper or lower decision criterion⁶. Although the paper does not discuss crossmodal interaction effects, the model could easily be applied to such data. A comparison of this model, which is not based on the concept of a temporal window of integration, with the extended TWIN model introduced here would be of interest but is

⁶For details, see Schwarz (2006).

beyond the scope of this note. Given that the TWIN model can be seen as an extended version of race models, such a comparison could be especially revealing given the recent equivalence results between diffusion and race models in (Jones & Dzhafarov, 2014).

Acknowledgments

This work was supported by DFG (German Science Foundation) Grant DI 506-12-1 to A.D. and DFG-SFB/TRR-31 (Project B4) to H.C. We thank Jim Townsend and the anonymous reviewers for helpful comments.

Abbreviations

RT	reaction time
RMI	race model inequality
SFE	statistical facilitation effect
SJ	simultaneity judgment
SOA	stimulus onset asynchrony
TOJ	temporal order judgment
TWIN	time-window-of-integration (model)

Appendix

Computation of psychometric function $\Psi(\tau)$. First, we have to evaluate the following integrals for positive and negative values of τ :

$$\begin{aligned} p_1 &= \Pr(A + \tau < V < A + \tau + \omega) &&= \int_0^\infty \{F_V(a + \tau + \omega) - F_V(a + \tau)\} dF_A(a), \\ p_2 &= \Pr(V < A + \tau < V + \omega) &&= \int_0^\infty \{F_A(v + \omega - \tau) - F_A(v - \tau)\} dF_V(v), \\ p_3 &= \Pr(V < A + \tau - \omega) &&= \int_0^\infty F_V(a + \tau - \omega) dF_A(a). \end{aligned}$$

For $\tau < 0$,

$$p_1 = \begin{cases} \frac{\lambda_V}{\lambda_V + \lambda_A} \{\exp[\lambda_A(\tau + \omega)] - \exp[\lambda_A\tau]\} & \text{if } \tau + \omega < 0; \\ \frac{\lambda_A}{\lambda_V + \lambda_A} \{1 - \exp[-\lambda_V(\tau + \omega)]\} + \frac{\lambda_V}{\lambda_V + \lambda_A} \{1 - \exp[-\lambda_A\tau]\} & \text{if } \tau < 0 < \tau + \omega; \end{cases}$$

and

$$p_2 = \frac{\lambda_V}{\lambda_V + \lambda_A} \{\exp[\lambda_A\tau] - \exp[-\lambda_A(\tau - \omega)]\};$$

and

$$p_3 = \frac{\lambda_V}{\lambda_V + \lambda_A} \exp[-\lambda_A(\tau - \omega)].$$

For $\tau > 0$,

$$p_1 = \frac{\lambda_A}{\lambda_V + \lambda_A} \{\exp[-\lambda_V\tau] - \exp[-\lambda_V(\tau + \omega)]\},$$

$$p_2 = \begin{cases} \frac{\lambda_A}{\lambda_V + \lambda_A} \{\exp[-\lambda_V(\tau - \omega)] - \exp[-\lambda_V\tau]\} & \text{if } \tau > \omega, \\ \frac{\lambda_A}{\lambda_V + \lambda_A} \{1 - \exp[-\lambda_V\tau]\} + \frac{\lambda_V}{\lambda_V + \lambda_A} \{1 - \exp[-\lambda_A(\omega - \tau)]\} & \text{if } \tau < \omega, \end{cases}$$

and

$$p_3 = \begin{cases} \frac{\lambda_V}{\lambda_V + \lambda_A} \exp[-\lambda_A(\omega - \tau)] & \text{if } \tau < \omega, \\ 1 - \frac{\lambda_A}{\lambda_V + \lambda_A} \exp[-\lambda_V(\tau - \omega)] & \text{if } \tau > \omega. \end{cases}$$

For the computation of the TOJ probabilities under the TWIN model, the cases of $\tau < 0$ and $\tau > 0$ must be considered separately, denoted here as $\Psi^-(\tau)$ and $\Psi^+(\tau)$:

$$\Psi^-(\tau) = \begin{cases} \frac{\lambda_V}{\lambda_V + \lambda_A} \{ \exp[\lambda_A(\tau - \omega)] (1 + \beta(-1 + \exp[2\lambda_A\omega])) \} & \text{if } \tau + \omega < 0; \\ \frac{1}{\lambda_V + \lambda_A} \{ \exp[\lambda_A(\tau - \omega)] \lambda_V + \beta(\lambda_A(1 - \exp[-\lambda_V(\omega + \tau)]) + \lambda_V(1 - \exp[\lambda_A(-\omega + \tau)])) \} & \text{if } 0 < \tau + \omega; \end{cases} \quad (3)$$

and

$$\Psi^+(\tau) = \begin{cases} 1 - \frac{\lambda_A}{\lambda_V + \lambda_A} \{ \exp[-\lambda_V(\omega + \tau)] [\beta - (\beta - 1) \exp[2\lambda_V\omega]] \} & \text{if } \tau > \omega; \\ \frac{1}{\lambda_V + \lambda_A} \{ \exp[\lambda_A(-\omega + \tau)] \lambda_V + \beta[\lambda_A(1 - \exp[-\lambda_V(\omega + \tau)]) + \lambda_V(1 - \exp[\lambda_A(-\omega + \tau)])] \} & \text{if } \tau < \omega; \end{cases} \quad (4)$$

Computation of expected reaction times in the crossmodal and unimodal conditions. We have, from Equation 1 in the main text,

$$E[RT_{V,A,\tau}] = E[\min(V, A + \tau)] + \mu - P(I) \times \Delta.$$

This becomes, after inserting the exponential distributions in $E[\min(V, A + \tau)]$,

$$E[RT_{V,A,\tau}] = \frac{1}{\lambda_V} - \exp[-\lambda_V\tau] \left(\frac{1}{\lambda_V} - \frac{1}{\lambda_V + \lambda_A} \right) + \mu - P(I)\Delta.$$

For the unimodal conditions, we get

$$E[RT_V] = \frac{1}{\lambda_V} + \mu \quad \text{and} \quad E[RT_A] = \frac{1}{\lambda_A} + \mu.$$

References

- Alais, D., & Carlile, S. (2005, Feb). Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound. Proceedings of the National Academy of Science U.S.A., 102(6), 2244–2247.
- Busemeyer, J., & Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision-making in an uncertain environment. Psychological Review, 100(3), 432–459.
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. Attention, Perception & Psychophysics, 75(5), 790–811.
- Colonus, H. (1990). Possibly dependent probability summation of reaction time. Journal of Mathematical Psychology, 34, 253–275.
- Colonus, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. J Cogn Neurosci, 16, 1000–1009.
- Colonus, H., & Diederich, A. (2006). Race model inequality: Interpreting a geometric measure of the amount of violation. Psychological Review, 113(1), 148–154.
- Corneil, B., van Wanrooij, M., Munoz, D., & van Opstal, A. (2002). Auditory-visual interactions subserving goal-directed saccades in a complex scene. Journal Neurophysiology, 88(7), 438–2002.
- Davison, A., & Hinkley, D. (1997). Bootstrap methods and their applications. New York: NY: Cambridge University Press.
- de Boer-Schellekens, L., & Vroomen, J. (2014). Multisensory integration compensates loss of sensitivity of visual temporal order in the elderly. Experimental Brain Research, 232, 253–262.
- Diederich, A. (1992). Intersensory facilitation: Race, superposition, and diffusion models for reaction time to multiple stimuli. Verlag Peter Lang.
- Diederich, A. (1994). A diffusion model for intersensory facilitation of reaction time. In G. Fischer & D. Laming (Eds.), Contributions to mathematical psychology,

- psychometrics, and methodology (pp. 207–220). New York: Springer Verlag.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. Journal of Mathematical Psychology, 39(2), 197–215.
- Diederich, A., & Colonius, H. (2004). Handbook of multisensory processes. In G. Calvert, C. Spence, & E. Stein B. (Eds.), (pp. 395–408). Cambridge: MIT Press.
- Diederich, A., & Colonius, H. (2007a). Modeling spatial effects in visual-tactile reaction time. Perception and Psychophysics, 69(1), 56–67.
- Diederich, A., & Colonius, H. (2007b). Why two "distractors" are better than one: Modeling the effect on non-target auditory and tactile stimuli on visual saccadic reaction time. Experimental Brain Research, 179, 43–54.
- Diederich, A., & Colonius, H. (2008a). Crossmodal interaction in saccadic reaction time: Separating multisensory from warning effects in the time window of integration model. Experimental Brain Research, 186(1), 1–22.
- Diederich, A., & Colonius, H. (2008b). When a high-intensity "distractor" is better than a low-intensity one: Modeling the effect of an auditory or tactile nontarget stimulus on visual saccadic reaction time. Brain Research, 1242, 219–230.
- Dixon, N., & Spitz, L. (1980). The detection of auditory visual desynchrony. Perception, 9, 719–721.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. Nature Neuroscience, 7, 773–778.
- García-Pérez, M., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal order judgment tasks: a quantitative model. Psychonomic Bulletin & Review, 19, 820–846.
- Gibbon, J., & Rutschmann, R. (1969). Temporal order judgment and reaction time. Science, 165, 413–415.
- Gondan, M. (2010). A permutation test for the race model inequality. Behavior Research Methods, 42(1), 23–28.

- Grice, G. (1968). Stimulus intensity and response evocation. Psychological Review, 75, 359–373.
- Harrar, V., & Harris, L. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. Experimental Brain Research, 186(4), 517–524.
- Harrar, V., Spence, C., & Harris, L. (2013). Does temporal adaptation affect multisensory integration? Multisensory Research Supplement, 26, 138.
- Hillock, A., Powers, A., & Wallace, M. (2011). Binding of sights and sounds: age-related changes in multisensory temporal processing. Neuropsychologia, 49, 461–467.
- Hillock-Dunn, A., & Wallace, M. (2012). Developmental changes in multisensory temporal binding window persist into adolescence. Developmental Science, 15(5), 688–696.
- Hollander, M., & Wolfe, D. (1999). Nonparametric statistical methods. New York: NY: John Wiley & Sons, Inc.
- Jáskowski, P. (1992). Temporal order judgment and reaction time for short and long stimuli. Psychological Research, 54, 141–145.
- Jáskowski, P., Jaroszyk, F., & Hojan-Jeziarska, D. (1990). Temporal-order judgements and reaction time for stimuli of different modalities. Psychological Research, 52, 35–38.
- Jones, M., & Dzhafarov, E. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. Psychological Review, 121(1), 1–32.
- Kandil, F., Diederich, A., & Colonius, H. (2014). Parameter recovery for the time-window-of-integration (twin) model of multisensory integration in focused attention. Journal of Vision, 14(11), 1–20.
- Koerding, K., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in multisensory perception. PLoS ONE, 2(9), e943.
- Kösem, A., & Van Wassenhove, V. (2013). Phase encoding of perceived events timing. Multisensory Research, 26(89).
- Laming, D. (1968). Information theory of choice-reaction times. New York: Academic

Press.

- Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. Frontiers in Psychology, 4.
- Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. PLoS ONE, 8(8), e71608. Retrieved from <http://dx.doi.org/10.1371/doi:10.1371/journal.pone.0071608>
- Mendonca, C., Escher, A., van de Par, S., & Colonius, H. (2014). Predicting auditory space calibration from recent multisensory experience. (unpublished manuscript)
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. Cognitive Psychology, 14, 247–279.
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention tasks. Perception & Psychophysics, 40, 331–343.
- Miller, J., & Schwarz, W. (2006). Dissociations between reaction times and temporal order judgments: a diffusion model approach. Journal of Experimental Psychology: Human Perception and Performance, 32, 394–412.
- Navarra, J., Hartcher-O'Brien, J., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. Proceedings of the National Academy of Science U.S.A., 106(23), 9169–9173.
- Neumann, O., Esselmann, U., & Klotz, W. (1993). Differential effects of visual-spatial attention on response latency and temporal-order judgment. Psychological Research, 56, 26–34.
- Neumann, O., & Niepel, M. (2004). Timing of perception and perception of time. In C. Kaernbach, E. Schroeger, & H. Mueller (Eds.), In psychophysics beyond sensation: Laws and invariants of human cognition (pp. 245–269). Mahwah, N.J.: Erlbaum.
- Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral systems states: a hypothesis. Naturwissenschaften, 77, 89–91.

- Powers, A., Hevey, M., & Wallace, M. (2012). Neural correlates of multisensory perceptual learning. Journal of Neuroscience, 32(18), 6263–6274.
- Powers, A., Hillock, A., & Wallace, M. (2009). Perceptual training narrows the temporal window of multisensory binding. Journal of Neuroscience, 29(39), 12265–12274.
- Raab, D. (1962). Statistical facilitation of simple reaction time. Transactions of the New York Academy of Sciences, 24, 574–590.
- Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59–108.
- Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: high-density electrical mapping of auditory–somatosensory integration. Autism Research, 3(5), 253–267.
- Rutschmann, J., & Link, R. (1964). Perception of temporal order of stimuli differing in sense mode and simple reaction time. Perceptual and Motor Skills, 18, 345–352.
- Sanford, A. (1971). Effects of changes in the intensity of white noise on simultaneity judgements and simple reaction time. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 23(A), 296–303.
- Sanford, A. (1974). Attention bias and the relation of perception lag to simple reaction time. Journal of Experimental Psychology, 102, 443–446.
- Schwarz, W. (1994). Diffusion, superposition, and the redundant-targets effect. Journal of Mathematical Psychology, 38, 504–520.
- Schwarz, W. (2006). On the relationship between the redundant signals effect and temporal order judgments: Parametric data and a new model. Journal of Experimental Psychology: Human Perception and Performance, 32(3), 558–573.
- Smith, P. (1990). A note on the distribution of response times for a random walk with gaussian increments. Journal of Mathematical Psychology, 34, 445–459.
- Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. Current Biology, 13, R519–R521.
- Sternberg, S., & Knoll, R. (1973). The perception of temporal order: Fundamental issues

- anda general model. In S. Kornblum (Ed.), Attention and performance iv (pp. 629–685). New York: Academic Press.
- Stevenson, R., Siemann, J., Schneider, B., Eberly, H., Woynaroski, T., Camarata, S., & Wallace, M. (2014). Multisensory temporal integratino in autism spectrum disorders. Journal of Neuroscience, 34(3), 691–697.
- Stevenson, R., Zemtsov, R., & Wallace, M. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. Journal of Experimental Psychology: Human Perception and Performance, 38(6), 1517–1529.
- Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. Perception & Psychophysics, 42, 224–239.
- van der Burg, E., Alais, D., & Cass, J. (2013). Rapid calibration to audiovisual asynchrony. Journal of Neuroscience, 33(37), 14633–14637.
- van Eijk, R., Kohlrausch, A., Juola, J., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. Perception & Psychophysics, 70, 955–968.
- van Eijk, R., Kohlrausch, A., Juola, J., & van de Par, S. (2009). Temporal interval discrimination thresholds depend on perceived synchrony for audio-visual pairs. Journal of Experimental Psychology: Human Perception and Performance, 35(4), 1254–1263.
- van Wanrooij, M., Bell, A., Munoz, D., & van Opstal, A. (2009). The effect of spatial-temporal audiovisual disparities on saccades in a complex scene. Experimental Brain Research, 198, 425–437.
- Van Wassenhove, V., Grant, K., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. Neuropsychologia, 45, 598–607.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. Attention, Perception & Psychophysics, 72(4), 871–884.

Wozny, D., & Shams, L. (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. Journal of Neuroscience, 31(12), 4607–4612.

Yarrow, K., Jahn, N., Durant, S., & Arnold, D. (2011). Shifts of criteria or neural timing? the assumptions underlying timing perception studies. Consciousness and Cognition, 20, 1518–1531.

Table 1

TWIN parameter estimates and confidence intervals for the bootstrapped time window width differences. Positive confidence limits indicate that ω_{RT} is larger than ω_{TOJ} , negative ones indicate the reverse order.

Participant	$1/\lambda_V$	$1/\lambda_A$	μ	Δ	β	ω_{RT}	ω_{TOJ}	99% CI
1	70.24	20.04	244.53	50.40	1.00	432.67	56.25	[372, 381]
2	102.57	43.85	222.02	53.91	0.68	591.26	154.47	[814, 922]
3	116.73	76.48	103.13	-80.34	0.73	609.44	202.51	[454, 501]
4	55.42	29.26	214.73	19.11	1.00	67.69	62.09	[12, 17]
5	107.76	48.95	172.31	51.60	1.00	32.96	87.32	[-69, -66]
6	51.29	28.42	219.41	16.95	1.00	105.32	61.96	[46, 50]
7	96.72	72.18	200.70	20.38	0.10	358.80	36.74	[349, 359]
8	112.35	92.28	130.92	-68.20	0.36	25.39	174.83	[-140, -134]
9	121.81	89.99	249.75	48.54	0.48	432.87	169.23	[1179, 1342]
10	57.14	20.00	217.08	49.61	0.84	487.04	56.21	[454, 501]
11	103.78	42.86	264.64	17.60	0.80	312.84	137.70	[217, 226]

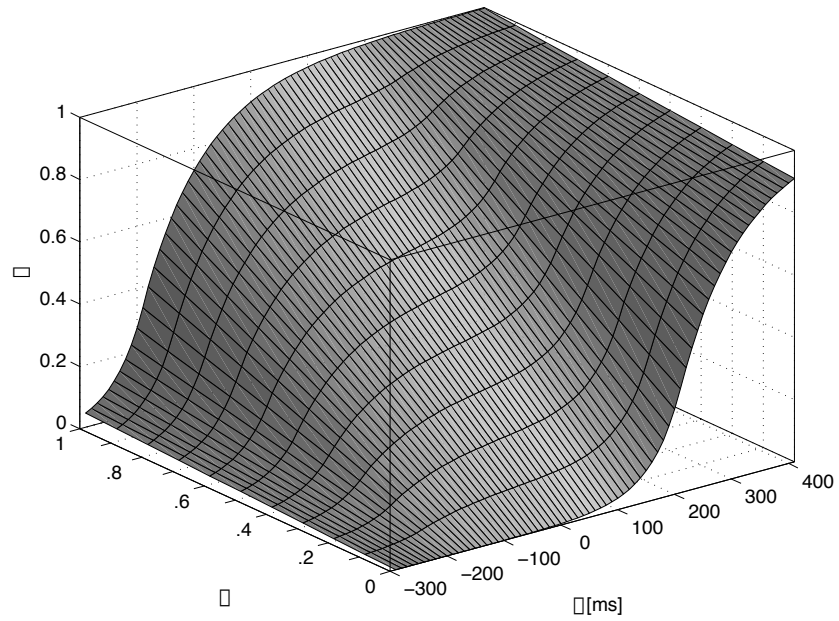


Figure 1. Probability of a “visual first” response as a function of SOA (τ) and response bias (β); other parameters are fixed: $\omega = 200$ [ms], $\lambda_V = 1/100$, $\lambda_A = 1/50$.

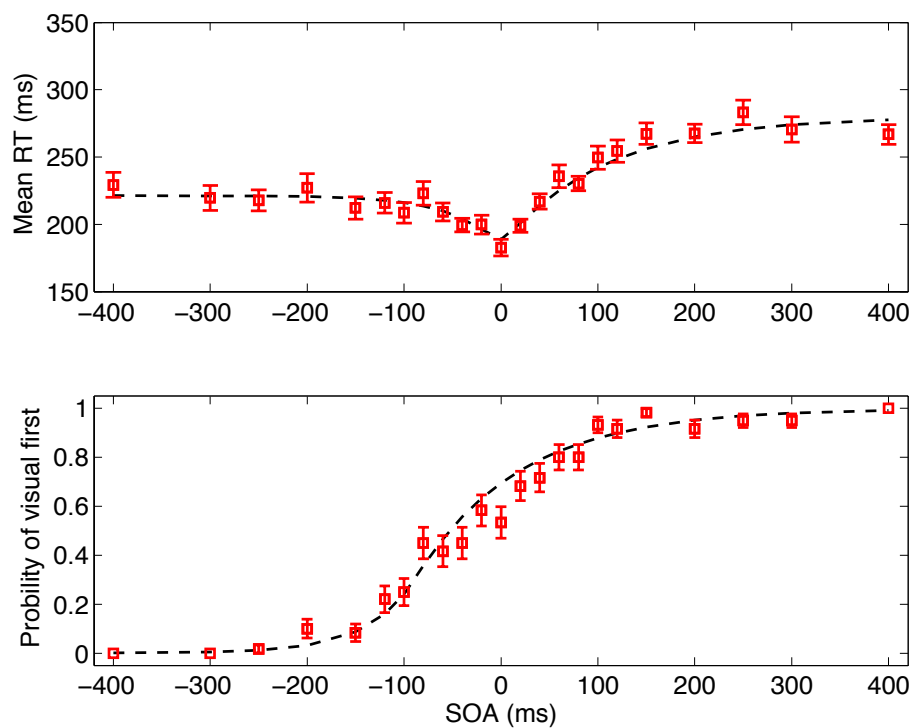


Figure 2. Example fit (\pm standard error) of mean bimodal RT (*upper panel*) and the psychometric function of judging “visual first” (*lower panel*) across the entire SOA range (participant 5). Rounded parameter estimates are $1/\lambda_V = 108$, $1/\lambda_A = 49$, $\mu = 172$, $\Delta = 52$, $\beta = 1.00$, $\omega_{RT} = 33$, and $\omega_{TOJ} = 87$.