Exp Brain Res
DOI 10.1007/s00221-011-2732-x

RESEARCH ARTICLE

Computing an optimal time window of audiovisual integration in focused attention tasks: illustrated by studies on effect of age and prior knowledge

Hans Colonius · Adele Diederich

Received: 10 March 2011 / Accepted: 12 May 2011

Abstract The concept of a "time window of integration" holds that information from different sensory modalities must not be perceived too far apart in time in order to be integrated into a multisensory perceptual event. Empirical estimates of window width differ widely, however, ranging from 40 to 600 ms depending on context and experimental paradigm. Searching for theoretical derivation of window width, Colonius and Diederich (*Front Integr Neurosci* 2010) developed a decision-theoretic framework using a decision rule that is based on the prior probability of a common source, the likelihood of temporal disparities between the unimodal signals, and the payoff for making right or wrong decisions. Here, this framework is extended to the *focused attention* task where subjects are asked to respond to signals from a target modality only. Evoking the framework of the time-window-of-integration (TWIN) model, an explicit expression for optimal window width is obtained. The approach is probed on two published focused attention studies. The first is a saccadic reaction time study assessing the efficiency with which multisensory integration varies as a function of aging. Although the window widths for young and older adults differ by nearly 200 ms, presumably due to their different peripheral processing speeds, neither of them deviates significantly from the optimal values. In the second study, head saccadic reactions times to a perfectly aligned audiovisual stimulus pair had been shown to depend on the prior probability of spatial alignment. Intriguingly, they reflected the magnitude of the time-window widths predicted by our decision-theoretic framework, i.e., a larger time window is associated with a higher prior probability.

Keywords Multisensory integration · Optimal time window of integration focused attention · Saccadic reaction time

Introduction

Evidence for multisensory integration is found in many different forms, notably as facilitation or inhibition of responses to a crossmodal stimulus set, compared to unimodal stimulation. In an orienting task, multisensory integration manifests in the speedup, or slowing, of eye and head mean saccadic reaction time. Besides facilitation or inhibition, crossmodal stimulation may result in no discernable effect beyond what would be expected from unimodal stimulation alone. This has sometimes been referred to as a situation where the crossmodal stimulus combination "falls outside of a spatiotemporal window of integration" (e.g., Meredith 2002). The *time-window hypothesis* holds that information from different sensory modalities must not be perceived too far apart in time, so that integration into a multisensory (perceptual) unit may occur. Empirical estimates of window width differ widely, however, ranging from 40 to 600 ms, depending on context and experimental paradigm (cf. Diederich and Colonius 2011, for a review).

H. Colonius (✉)
Department of Psychology, Oldenburg University, Oldenburg, Germany
e-mail: hans.colonius@uni-oldenburg.de

A. Diederich
School of Humanities and Social Sciences, Jacobs University Bremen, 28725 Bremen, Germany
e-mail: a.diederich@jacobs-university.de

H. Colonius · A. Diederich
Research Center Neurosensory Science, 26111 Oldenburg, Germany

Published online: 31 May 2011

A hitherto open question has been whether window width, rather than being estimated from empirical data, may also be determined on theoretic grounds. We recently suggested a decision-theoretic framework to compute optimal window width in a reaction time task with *redundant* targets where stimuli of either modality may be responded to (Colonius and Diederich 2010). Since multisensory integration studies involving *nontarget* modalities are abundant, here we derive time window optimality in a focused attention (FA) task where stimuli from the nontarget modality could be ignored. Moreover, we compare the optimal window estimates with the empirical results from two recent audiovisual FA studies where actual window width was presumably modulated (i) by participants' age and (ii) by the prior probability of the spatial distance between visual and auditory stimuli.

Next, to provide an appropriate framework for the derivation of an optimal time window of integration we describe the time-window-of-integration (TWIN) model (see Diederich and Colonius 2008a, for further details).

Time-window-of-integration model

A classic explanation for a speed-up of responses to crossmodal stimuli is that subjects are merely responding to the first stimulus detected. Taking the detection times to be random variables and adding some technical assumptions, observed reaction time is representable by the minimum of the reaction times to the, say, visual or auditory signal, leading to a purely statistical facilitation effect in response speed (Raab 1962). However, numerous studies have shown that this *race model* is not sufficient to explain the observed speedup in reaction time (e.g., Diederich and Colonius 1987; Giray and Ulrich 1993; Corneil and Munõz 1996; Van Wanrooij et al. 2009; for a review, see Diederich and Colonius 2004). Using the race model inequality (Miller 1982; Colonius and Diederich 2006) as a benchmark test, responses to bimodal stimuli have been found to be faster than predicted by statistical facilitation, in particular, when the stimuli were spatially aligned (Frens et al. 1995; Hughes et al. 1998; Harrington and Peck 1998). Note that although the race model test has been applied to data from FA tasks, this practice seems dubious given that the effect of a stimulus from the nontarget modality winning the race is not specified in the model. Moreover, the race model can not predict the inhibition often observed with spatially disparate stimuli.

Although the race model in its simple form cannot account for the empirical data, the basic concept of a race occurring at a very early stage of processing has considerable plausibility. Hence, the time-window-of-integration (TWIN) model postulates that a crossmodal stimulus triggers a race mechanism in the very early, peripheral sensory pathways. This first stage is followed by a compound stage of converging subprocesses that comprise neural integration of the input and preparation of a response (for details, see Diederich and Colonius 2011; Colonius and Diederich 2004). Note that 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 TWIN concerns the temporal configuration needed for multisensory integration to occur: Multisensory integration occurs only if the peripheral processes of the first stage all terminate within a given temporal interval, the "time window of integration" (TWIN assumption). 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 need not be sufficient, for crossmodal interaction to occur since the manifestation of interaction may also depend on many other aspects of the stimulus set, like spatial configuration of the stimuli (Frens et al. 1995), stimulus motion (Soto–Faraco et al. 2002), intra-modal grouping (Sanabria et al. 2005), semantic congruity (Molholm et al. 2004), or their synesthetic correspondence (Parise and Spence 2009). Note that the all-or-none assumption about the filter refers to a single trial only. On the average, the amount of crossmodal interaction [measured in ms] is weighted by the probability of interaction occurring, a term that does depend on SOA (cf. Eq. 3 below). Given the quantitative formulation of the TWIN model, this assumption is testable and, so far, our empirical tests seem to support it.

The amount of crossmodal interaction is shown as an increase or decrease of second stage processing time, but it is assumed not to depend on the stimulus onset asynchrony (SOA) of the stimuli. The basic feature of the TWIN framework is the priority of temporal proximity over any other type of proximity[1]. Rather than assuming a joint *spatiotemporal* window of integration permitting interaction to occur only for both spatially and temporally neighboring stimuli, the framework allows for crossmodal interaction to occur, for example, for spatially rather distant stimuli of different modalities as long as they fall within the time window. In its simplest form, the TWIN model framework is "agnostic" with respect to the exact nature of the integration mechanism occurring in the second stage. Nevertheless, it affords a number of quantitative predictions without making specific assumption about the probability distributions of the processing times. Several of these predictions have found empirical support in recent

[1] See Cohen (2011), for a related discussion of the role of time in cognitive neuroscience.

studies (cf. Diederich and Colonius 2007a, b, 2008a, b). In order to derive quantitative expressions for optimal time-window width, however, distributional assumptions for the first-stage processes are required.

Time-window-of-integration model: deriving quantitative predictions

For concreteness, assume an FA attention task with a visual target and an auditory nontarget stimulus. The race in the first stage is represented by two statistically independent, nonnegative random variables V and A denoting the peripheral processing times for the visual and the auditory stimulus, respectively. With τ as SOA value and ω as integration window width parameter, the requirement for multisensory integration to take place is that the nontarget stimulus A wins the race in the first stage "opening the time window of integration" such that the termination of the target peripheral process V falls into the window,

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

Here, a positive τ value indicates that the visual stimulus is presented before the auditory, and a negative τ value indicates the reverse presentation order. An intuitive interpretation of the definition of I is that the winning nontarget will keep the system in a state of heightened reactivity such that the upcoming target stimulus, if it falls into the time window, will trigger crossmodal interaction. For saccadic eye movements, this may in particular correspond to a gradual inhibition of fixation neurons (in superior colliculus) and/or omnipause neurons (in midline pontine brain stem). If the stimulus from the target modality is the winner of the race in the peripheral channels, second stage processing is initiated without any multisensory integration mechanism being involved. The rationale is that if the target is registered first there is "no reason" for the system to delay preparing the response. In the TWIN version for a "redundant signals task" (i.e., signals from any modality are targets), however, any stimulus can open the window (Colonius and Diederich 2010). These assumptions imply that, under identical stimulus conditions but different task instructions, crossmodal effects should be larger for the redundant than for the focused attention task. This also seems to be supported by data (Diederich and Colonius 2011).

Let S_1 and S_2 denote the random processing time in the first and second stage, respectively, so that total reaction time in the crossmodal condition becomes

$$RT_{VA} = S_1 + S_2. \tag{1}$$

Let $P(I)$ be the probability for the event I defined above. The expected reaction time in the bimodal condition can then be written as

$$\begin{aligned}\mathrm{E}[RT_{VA}] &= \mathrm{E}[S_1] + \mathrm{E}[S_2] \\ &= \mathrm{E}[S_1] + P(I)\cdot\mathrm{E}[S_2|I] + [1 - P(I)]\cdot\mathrm{E}[S_2|I^c],\end{aligned}$$

where $\mathrm{E}[S_2|I]$ and $\mathrm{E}[S_2|I^c]$ denote the expected second stage processing time conditioned on event I occurring or not occurring (I^c), respectively. Putting $\Delta \equiv \mathrm{E}[S_2|I^c] - \mathrm{E}[S_2|I]$, this becomes

$$\mathrm{E}[RT_{VA}] = \mathrm{E}[S_1] + \mathrm{E}[S_2|I^c] - P(I)\cdot\Delta. \tag{2}$$

That is, mean RT to bimodal stimuli is the sum of mean RT of the first-stage processing time, mean RT of the second stage processing when no interaction occurs, minus the term $P(I)\cdot\Delta$, which is a measure of the expected amount of intersensory interaction in the second stage with positive Δ values corresponding to facilitation, negative ones to inhibition.

Adding the exponential distribution assumption

Since in the following we only consider predictions of TWIN at the level of mean reaction times, no assumption about the distribution of second stage processing time, S_2, is needed, so we simply introduce a parameter μ denoting $\mathrm{E}[S_2|I^c]$.

In order to compute $P(I)$,we postulate exponential distributions for the peripheral processes, V and A, with expected values $1/\lambda_V$ and $1/\lambda_A$, respectively (λ_V, λ_A are positive-valued parameters characterizing the exponential distributions). The choice of exponentials is mainly for ease of computation and could be revised if the data suggests. Mean RT in the bimodal condition then is

$$\mathrm{E}[RT_{VA}] = \frac{1}{\lambda_V} + \mu - P(I)\cdot\Delta,$$

where $P(I)$ is a function of λ_V, λ_A, window width ω, and the SOA value τ (see "Appendix 1"). Note that in the FA task, first-stage duration is defined as the peripheral processing time for the (visual) target stimulus. Thus, mean RT in the unimodal condition is

$$\mathrm{E}[RT_V] = \frac{1}{\lambda_V} + \mu.$$

Crossmodal interaction (CI) is defined as the difference between mean RT to the unimodal and crossmodal stimuli, i.e.,

$$\mathrm{CI} \equiv \mathrm{E}[RT_V] - \mathrm{E}[RT_{VA}] = P(I)\cdot\Delta. \tag{3}$$

This represents the factorization of CI into a factor $P(I)$ only depending on stimulus parameters λ_V, λ_A, window width ω, and SOA τ on the one hand, and a factor Δ only depending on parameters characterizing the bimodal context, like the spatial separation between the stimuli, on the other. It is the basis for many of the empirical tests of the TWIN framework (cf. Diederich and Colonius 2011).

Toward an optimal time window

We now introduce the decision-theoretic framework that affords to determine an optimal time-window width. Usually, the width parameter ω is estimated in TWIN from the data through optimizing model fit with respect to some deviance criterion. This often yields interpretable results; for example, Diederich et al. (2008) found windows of different sizes for two different age groups which they attributed to a relative slowing of peripheral processing in the old age group (see "The Diederich et al. 2008 study" below). Deriving an optimal parameter value for ω, one can examine whether, or to what degree, subjects in a given context are able to adapt their window width to this optimal value.

The basic decision situation

It has been suggested that integrating crossmodal information always involves a –possibly implicit– decision about whether or not two (or more) sensory cues originate from the same event, i.e., have a common cause (e.g., Körding et al. 2007; Stein and Meredith 1993). For example, in a predator–prey situation, when the potential prey perceives a sudden movement in the dark, it may be vital to recognize whether this is caused by a predator or a harmless wind gust. If visual information is accompanied by some vocalization from a similar direction, it may be adequate to respond to the potential threat by assuming that the visual and auditory information are caused by the same source, i.e., to perform multisensory integration leading to a speeded escape reaction. On the other hand, in such a rich dynamic environment it may also be disadvantageous, e.g., leading to a depletion of resources, or even hazardous, to routinely combine information associated with sensory events which—in reality—may be entirely independent and unrelated.

The basic decision situation is presented in a schematic manner by the following "payoff matrix" (Table 1). It defines the gain (or cost) function U associated with the *states of nature* (C) and the *action* (I) of audiovisual integration:

Table 1 Payoff matrix for the basic decision situation

	Integration ($I = 1$)	No integration ($I = 0$)
Common source ($C = 1$)	U_{11}	U_{10}
Separate sources ($C = 2$)	U_{21}	U_{20}

Variable C indicates whether visual and auditory stimulus information are generated by a common source ($C = 1$), i.e., an *audiovisual event*, or by two separate sources ($C = 2$), i.e., auditory and visual stimuli are unrelated to each other. Variable I indicates whether or not integration occurs ($I = 1$ or $I = 0$, respectively). The values U_{11} and U_{20} correspond to correct decisions and will in general be assumed to be positive numbers, while U_{21} and U_{10}, corresponding to incorrect decisions, will be negative. The organism's task is to balance these costs and benefits of multisensory integration by an appropriate optimizing strategy (cf. Körding et al. 2007).

Deriving an optimal decision rule

In Colonius and Diederich (2010), the following decision-theoretic approach toward finding the optimal time window has been proposed. Let us assume that, at a given point in time, *a-priori* probabilities for the events $\{C = i\}_{i=1,2}$ exist, with $P\ (C = 1) = 1 - \mathrm{P}(C = 2)$. In general, an optimal strategy may involve many different aspects of the empirical situation, like spatial and temporal contiguity. For example, Sato et al. (2007) take into account both spatial and temporal conditions simulating performance in an audiovisual localization task. Here, the simplifying assumption is that the temporal disparity between the "arrival times" of the unimodal signals is the *only* perceptual evidence utilized by the organism. Thus, computation of the optimal time window will be based on the prior probability of common cause and the likelihood of temporal disparities between the unimodal signals.

Let the temporal disparity between the arrival times of the unimodal signals be denoted by the random variable T, the arrival time difference (ATD). For a realization t of T, we define the *likelihood function* $f(t|C)$, where f denotes the probability mass function or, if it exists, the density function of T given C. Using Bayes' rule, we immediately have the *posterior* probability of a common cause given the occurrence of an arrival time difference t,

$$\mathrm{P}(C = 1|t) = \frac{f(t|C = 1)P(C = 1)}{f(t|C = 1)\mathrm{P}(C = 1) + f(t|C = 2)\mathrm{P}(C = 2)}.$$

Decision rule: Maximize the expected value of U. On each trial, in order to maximize the expected value $\mathrm{E}[U]$ of function U in the payoff matrix (Table 1), the decision-making mechanism is to choose that action alternative (i.e., to integrate or not) which contributes, on the average, more to $\mathrm{E}[U]$ than the other action alternative. Introducing the *likelihood ratio* function

$$\mathrm{L}(t) = f(t|C = 1)/f(t|C = 2),$$

results in the following decision rule (cf. Colonius and Diederich 2010):

$$\text{``}\mathit{If}\quad L(t) > \frac{P(C=2)}{P(C=1)} \times \frac{U_{20}-U_{21}}{U_{11}-U_{10}}, \tag{4}$$
integrate, otherwise do not integrate.''

This decision rule implicitly defines a window that is optimal in the sense of maximizing E[U].

> *The* optimal time window *is the set of all values of arrival time differences* $\{T = t\}$ *satisfying the inequality in the above decision rule* (4).

The effect of the prior probability for a common cause on the time window is immediately predicted from this decision rule: Keeping the U-values constant, the expression on the right of Inequality (4) will decrease as $P(C = 1)$ increases, implying an extension of the time window. Empirical support for this –intuitively plausible– prediction is reported below ("The van Wanrooij et al. 2010 study").

Computing an optimal window in the focused attention (FA) task

Applying the decision rule introduced above requires a specification of the likelihood function of the arrival time difference T under either hypothesis, $f(t|C = i)$, $\quad i = 1, 2$. Since T can take on either positive or negative values, the arrival order of the visual and auditory information is also available to the decision mechanism. Note, however, that we do not presuppose the existence of some high-level decision-making entity contemplating different action alternatives. It is only assumed that observed (average) behavior can be assessed as being consistent or inconsistent with an optimal strategy with respect to the time-window width.

For two separate sources, we assume a uniform law,

$$f(t|C=2) = \begin{cases} 1/(t_1 - t_0) & \text{if } t_0 < t < t_1, \\ 0 & \text{otherwise,} \end{cases}. \tag{5}$$

Here, t_0, t_1 are real numbers defining the *observation interval*, that is, the interval of time limiting all possible ATDs due to the construction of a trial by the experimenter. Thus, under two separate sources any arrival time difference is assumed to occur with the same likelihood within the observation interval (t_0, t_1).

For a single source, we postulate that the likelihood function is induced by the distributions of the peripheral processing times V and A, i.e., T has the same distribution as $V - A$. It is important to keep in mind that this is an additional assumption not following from the decision framework. It seems plausible, however, given that in a typical environment visual and auditory information deriving from a common source should occur more or less simultaneously.

Given the independent exponential distribution assumption for V and A in TWIN, the distribution of arrival time differences under a common source, $V - A$, can be shown to be an *asymmetric-Laplace distribution* (see "Appendices 1, 2"):

$$f(t|C=1) = \frac{\lambda_V \lambda_A}{\lambda_V + \lambda_A} \times \begin{cases} \exp(-\lambda_V t) & \text{if } t \geq 0, \\ \exp(\lambda_A t) & \text{if } t < 0. \end{cases} \tag{6}$$

Here λ_V and λ_A are the intensity parameters of the visual and auditory exponential distributions, respectively. Note that the asymmetry derives from the asymmetry of the role of the modalities in FA tasks (target vs. non-target).

For $t_0 \leq t \leq t_1$, the likelihood ratio becomes

$$\begin{aligned} L(t) &= f(t|C=1)/f(t|C=2) \\ &= \frac{\lambda_V \lambda_A}{\lambda_V + \lambda_A}(t_1 - t_0) \\ &\quad \times \begin{cases} \exp(-\lambda_V t) & \text{if } t \in (t_0, t_1) \cap [0, t_1), \\ \exp(\lambda_A t) & \text{if } t \in (t_0, t_1) \cap (t_0, 0] \end{cases} \end{aligned} \tag{7}$$

Figure 1 illustrates the likelihood ratio as a function of t under specified parameter values.

Note that for t outside of the observation interval the likelihood ratio remains undefined. To simplify the exposition in the following, the ratio of utility differences in Eq. 4 will be set equal to one. Thus, according to the optimal decision rule, audiovisual integration should be performed if and only if

$$L(t) > \frac{1-p}{p},$$

with $p \equiv \mathrm{P}(C = 1)$.

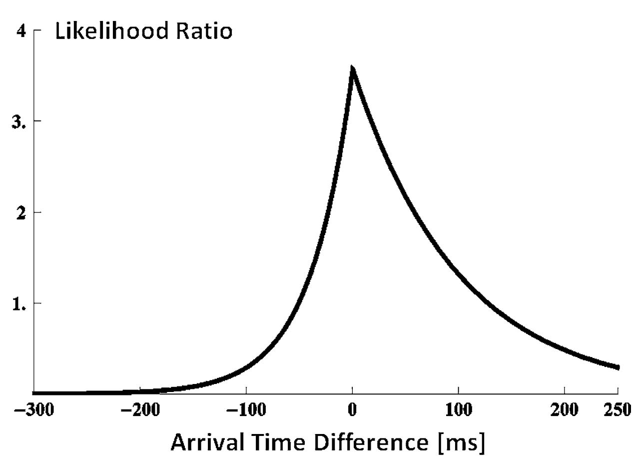

Fig. 1 Likelihood ratio function for asymmetric-Laplace/uniform densities example. Parameter values are $\lambda_V = .01$, $\lambda_A = .025$, $t_1 = 250$, $t_0 = -300$; the function is defined for $t \in (t_0, t_1)$

Inserting the expression for $L(t)$ from Eq. 7 and solving for t yields the following *optimal* time window for $t \in (t_0, t_1)$:

$$\left\{ t \middle| \frac{1}{\lambda_A} \log\left[\frac{(\lambda_V+\lambda_A)(1-p)}{\lambda_V\lambda_A(t_1-t_0)p}\right] \le t \le \frac{1}{\lambda_V}\log\left[\frac{\lambda_V\lambda_A(t_1-t_0)p}{(\lambda_V+\lambda_A)(1-p)}\right] \right\} \quad (8)$$

provided that

$$\frac{(\lambda_V+\lambda_A)(1-p)}{\lambda_V\lambda_A(t_1-t_0)p} \le 1. \quad (9)$$

This latter condition guarantees that the left side of the interval is non-positive and the right side is non-negative. For the width of the optimal time window, we get immediately

$$\omega_{opt} = \left(\frac{1}{\lambda_V} + \frac{1}{\lambda_A}\right) \log\left[\frac{\lambda_V\lambda_A(t_1-t_0)}{\lambda_V+\lambda_A}\frac{p}{1-p}\right] \quad (10)$$

This is obviously an increasing function of the prior odds $p/(1-p)$ and of the observation interval (t_0,t_1). Figure 2 shows the corresponding upper and lower limits of the optimal time window as a function of the prior probability of a common source, $p = P(C = 1)$. Increasing $P(C = 1)$ leads to a widening of the time window, in this case approaching infinity in a nonlinear fashion. Moreover, the optimal time window disappears for values of the prior below a certain positive threshold value. Although the exact threshold value depends on the experimental context (i.e., $t_1 - t_0$) and may get closer to zero, this prediction provides a potentially strong model test: for a small enough value of $P(C = 1)$ there should be no multisensory integration effect at all.

Empirical versus optimal time-window width

Once the TWIN model has been fitted to a set of data, the resulting parameters can be utilized to compute the optimal window width according to Eq. 10, and this optimal value can then be compared to the “observed” (i.e., fitted) window width. This approach will be illustrated in the next section with a study exploring the effect of age in an audiovisual orienting task.

Notably, even without fitting a parametric model for the time window, the decision rule introduced here (Eq. 4) lends itself to a number of empirical tests by (i) manipulating prior probability of a single source, (ii) the likelihood functions for the arrival time differences, or (iii) the utilities associated with the different outcomes listed in the payoff matrix in Table 1 (see also Colonius and Diederich 2010). For prior probabilities, this is illustrated below with data from a study of head saccades (van Wanrooij et al. 2010) with different priors for a single source.

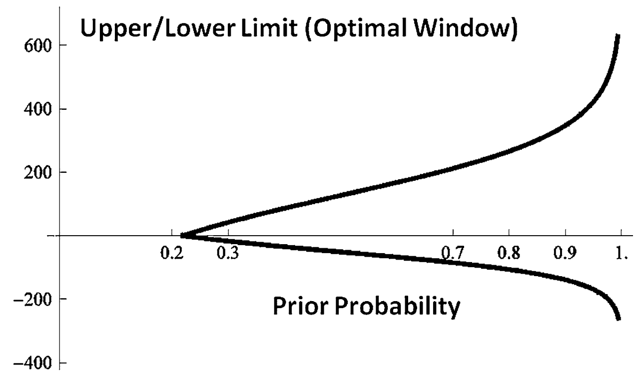

Fig. 2 Upper and lower limits of the optimal time window as a function of the prior probability for a common source $P(C = 1)$ under the asymmetric-Laplace/uniform likelihood functions of Figure. Optimal window width (upper minus lower limit) increases nonlinearly with prior probability

The Diederich et al. (2008) study

A number of studies assessing the efficiency with which multisensory integration occurs as a function of aging have found the benefit of multisensory signals to perception to be greater in older than younger adults (e.g., Peiffer et al. 2007; Laurienti et al. 2006; but see Hugenschmidt et al. 2009). For an example in response speed effects, we consider the study by Diederich et al. (2008) measuring saccadic RT of aged (65–75 years) and young individuals (20–22 years) to the onset of a visual target stimulus with and without an accessory auditory stimulus occurring ipsi- or contralateral to the target (FA task). The response time pattern for both groups was similar: mean RT to bimodal stimuli was generally shorter than to unimodal stimuli, and mean bimodal RT was shorter when the auditory accessory was presented ipsilateral rather than contralateral to the target. The elderly participants were considerably slower than the younger participants under all conditions but showed a greater multisensory enhancement, that is, they seemed to benefit more from bimodal stimulus presentation relative to unimodal presentation.

Fitting the TWIN model to the data from both age groups separately, the resulting numerical parameter values reflected the slowing of the peripheral sensory processing in the elderly ($1/\lambda_V = 84$ ms for visual and $1/\lambda_A = 98$ ms for auditory stimuli compared to 48 ms and 18 ms for the younger group, respectively) but, importantly, the elderly seemed to have a larger time window of integration ($\omega = 450$ ms vs. 275 ms). Note that the slowing of the peripheral processes makes the arrival time for visual and auditory sensory information more variable and thereby diminishes the likelihood of them terminating within the time window. It seems, however, that a broader temporal window could only partially compensate for this since the older group was nevertheless slower in the bimodal conditions than the younger one[2].

This raises the question of whether or not the elderly in this study may have an enlarged time window that is close to the optimal value. Computing the optimal window is straightforward: we first insert the estimated λ parameter values into Eq. 10. For the value of the prior probability of a single source, we choose $p = 0.5$ since ipsi- and contralateral configurations were presented with equal probability. In doing so ,we equate ipsilateral presentations with a single source ($C = 1$) and contralateral presentations with two separate sources ($C = 2$). The remaining parameter to be determined is the observation interval length, $t_1 - t_0$. Given the temporal structure of a trial in the experiment (foreperiod of up to 1,200 ms and maximal stimulus duration of 500 ms),we set $t_1 - t_0 = 1{,}700$ ms. With these parameter values inserted in Eq. 10, the optimal window width becomes 407 ms for the older group and 214 ms for the young adults group. Both values are a bit below the estimated widths (450 ms and 275 ms, respectively) but, given the variability in the parameter estimation process that we typically observe, this does not lend strong support to the hypothesis that either group deviates much from the optimal integration window size.

This computation also neglects the possibility of the utility values (costs/ benefits) influencing the decision process, as the ratio of utility values was set equal to one in the decision rule underlying the optimal window width. The fact that a large integration window may come at a certain cost is nicely demonstrated in a recent study by Setti at al. (2011). When information coming from different sources is erroneously combined, this can result in distractibility and inefficient processing of the relevant stimulus in older adults possibly rendering older adults more accident prone (Poliakoff et al. 2006). In their study, Setti at al. investigated whether the causes of falling, a significant problem for older persons, are related to audio–visual integration ability. They measured susceptibility to the sound-induced flash illusion (Shams et al. 2000) for three groups, fall-prone, non-fall prone older adults, and young adults. The flash illusion is induced when two auditory stimuli (beeps) presented with a single brief visual stimulus (flash) results in the perception of two visual 'flashes'. It occurs when the two sensory inputs are integrated due to their temporal proximity (Shams et al. 2002) but spatial proximity does not seem to play a role (Innes–Brown and Crewther 2009). Specifically, Setti et al. found that older adults with a history of falling integrated audio–visual stimuli over a longer delay between the onset of cross-sensory stimulations than either older adults without a history of falling or younger adults. In particular, fall-prone older adults experienced higher overall rates of the sound-induced flash illusion at longer SOAs (from 110 to 270 ms) than non-fallers. Setti at al. (ibid, p. 382) speculate that there may be either an indirect or a direct effect of a wider temporal window of integration on balance maintenance and posture control in fall-prone older adults.

The van Wanrooij et al. (2010) study

In a recent FA study, Van Wanrooij and colleagues (Van Wanrooij et al. 2010) measured human head saccades toward a visual target (50 ms flash) which, in the bimodal condition, was accompanied by a synchronous (white noise) sound (50 ms, 65 dB). Visual (v) and auditory (a) stimuli were presented either spatially aligned or vertically disparate in the midsagittal plane at 10 possible locations: ±15, 20, 25, 30 or 35° in elevation (with 0° at straight ahead). It is assumed that, in such a reduced laboratory situation with simple visual and auditory stimuli, spatial contiguity is the main determinant of perceiving visual and auditory information as a common crossmodal event, given a small enough arrival time difference. This premise is supported by the observation that facilitation of (saccadic) RT is maximal when visual and auditory stimuli appear at the same position in space and that it decreases, or even turns into inhibition, when spatial distance increases (Frens et al. 1995; Corneil and Munoz 1996; Colonius and Arndt 2001; Whitchurch and Takahashi 2006).

Consistent with those previous results, van Wanrooij et al. found (i) responses to spatially aligned audiovisual stimuli were, on average, faster than unimodal responses, and (ii) spatial disparity systematically delayed the audiovisual responses, by up to about 38 ms, when the auditory nontarget was presented in the hemifield opposite to the visual target. Importantly, the bimodal responses were recorded under three different conditions of spatial disparity distribution, collected in separate experiments. In the first experiment, the visual stimuli were always accompanied by spatially (and temporally) aligned sounds (*av-100/0*). In the second, the visual stimuli were aligned with the sound in 50% of the trials while the others had a spatial disparity of more than 45° (*av-50/50*). In the third, only 10% of the trials contained spatially aligned stimulus pairs while the others had spatial disparities between 5° and 75° (*av-10/90*). Although the actual prior probabilities for spatial alignment (thus, for a common audiovisual source) utilized by the subjects are not observable, the only plausible ordering of the prior probabilities is

[2] What remains unclear is whether the slowing is only due to an age-related decrement in peripheral processing or also in a more centrally located processing stage where integration of visual and auditory input is taking place.

$p_{av-100/0} > p_{av-50/50} > p_{av-10/90}$.

According to the decision rule introduced in Eq. 4, this ordering implies a corresponding ordering of the time-window width such that –under perfect spatial alignment– facilitation of reaction time should reflect the ordering. This is exactly what Van Wanrooij and colleagues found (see Fig. 3, redrawn from their data). Reaction times in conditions *av-50/50* and *av-10/90* were larger than in *av-100/0* for 9 out of 14 cases ($p < .001$) and the ordering was as predicted in all but one case.

Discussion

Several research groups have suggested that multisensory integration more or less closely follows rules based on optimal (Bayesian) inference procedures (e.g., Anastasio et al. 2000; Ernst and Banks 2002; Hillis et al. 2002; Battaglia et al. 2003; Alais and Burr 2004; Colonius and Diederich 2004b; Wallace et al. 2004; Shams et al. 2005; Sato et al. 2007; Ernst 2007; Beierholm et al. 2008; Roach et al. 2009; Di Luca et al. 2009; Wozny et al. 2010; see Ernst 2005, for a review). None of these studies, however, has considered determining an optimal temporal window of integration.

The notion of a time window of integration has become a widely accepted concept in multisensory research: crossmodal information falling within this window is (highly likely to be) integrated, whereas information falling outside is not (e.g., Meredith 2002; Colonius and Diederich 2004a; Powers 3rd et al. 2009; Lewkowicz 1996). Thus, integrating crossmodal information always implies a decision about whether or not two (or more) sensory cues originate from the same event, i.e., have a common cause. In Colonius and Diederich (2010), we have developed a decision-theoretic framework using a decision rule maximizing expected gain that is based on the prior probability of a common source and the likelihood of temporal disparities between the unimodal signals. Assuming the redundant signals paradigm, where subjects may respond to signals from any of the modalities, we could derive an explicit expression for the temporal window of *optimal* width.

Here, we have extended our previous approach to the important case of a focused attention (FA) task where subjects are asked to respond to signals from a target modality only. Again, the computation of an optimal window width requires explicit assumptions about the likelihood function for the arrival time difference given a common source for the audiovisual crossmodal information. However, since there is no longer a complete

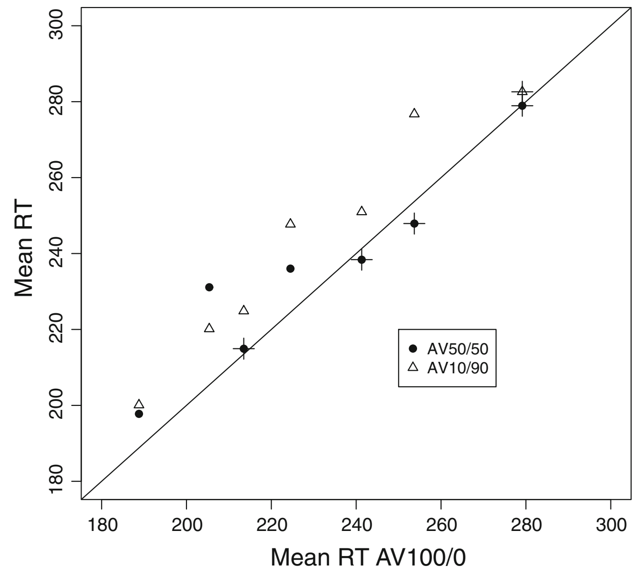

Fig. 3 Mean saccadic reaction time to spatially aligned audiovisual stimuli in the 50% alignment (*black dots*) and 10% alignment (*triangles*) conditions against the mean in the 100% alignment condition for 7 subjects in Van Wanrooij et al. (2010). Only one subject exhibited faster reactions in the 10% condition than in the 50% condition. Crossed-out data points were not significantly different from the 100% condition

symmetry in the task with respect to the modalities, both positive and negative arrival time differences may occur depending on whether the target or the nontarget modality process is terminated first. Evoking the framework of the TWIN model with exponentially distributed peripheral processing times we demonstrate that this implies an asymmetric-Laplace likelihood function, which is completely determined by the (intensity) parameters of the peripheral processing times.

Next, we have illustrated the empirical usefulness of our approach with two published FA studies. For the saccadic reaction time data from Diederich et al. (2008), we computed the optimal time window widths both for the younger and older adults group and compared them to the estimated widths. It turned out that, although the window widths for both groups differ by nearly 200 ms, presumably due to their different peripheral processing speeds, neither of them deviates significantly from the optimal values. The second study (van Wanrooij et al. 2010) is the first to systematically vary the prior probability of a single source for an audiovisual stimulus configuration. Intriguingly, mean (head saccadic) reaction times to a perfectly aligned audiovisual stimulus pair depended on the prior probability of spatial alignment, thus reflecting the order of the time-window widths predicted by our decision-theoretic framework (i.e., larger time window for higher prior probability, cf. Fig. 3).

While these empirical results are encouraging, the optimal time window approach suggested here could be tested further. The next step would be to modify systematically the likelihood functions of arrival time differences. Although arrival time differences are not observable, one can indirectly manipulate their likelihood distribution by changing the interstimulus interval (ISI): large ISI values should generate, on average, large arrival time differences. Another aspect that arose in the context of the study of age effects is the role of changing payoffs. Increasing the time window beyond a certain value may result in increasing costs (like in the fall-prone elderly group). More generally, manipulating the gains and costs for integrating visual and auditory information when they derive from a common source, and/or decreasing the costs when they don't, should lead to a larger window of integration and, thus, to shorter average RTs. In the laboratory, this can be achieved in the above setting through appropriate instruction, using different response deadlines and reward settings.

Another important aspect of the time-window hypothesis that should be investigated concerns the plasticity of its width. It is not yet entirely clear how much stimulus exposition is needed to establish, e.g., the prior probability of a common source, and how quickly changes in the experimental conditions will affect the setting of the time-window width. Recent results on the perception of audiovisual simultaneity suggest a high degree of flexibility in multisensory temporal processing (Lewkowicz 2010; Roseboom et al. 2009; Powers et al. 2009; Vroomen et al. 2004; Keetels and Vroomen 2007; Vroomen and Keetels 2010).

Finally, the notion of stimulus information coming from a "common source" should be given further scrutiny in our decision-making context since the prior probability parameter (p) depends on it[3]. In the two empirical studies considered here, "common source" was defined by spatial contiguity alone. Obviously, under richer environmental conditions the decision to treat crossmodal information as coming from a common source may depend on many other stimulus dimensions (for a recent review of empirical results, see Spence 2007). The more general concept of crossmodal binding is one of the central issue of multisensory integration research including attempts to theoretically define "audiovisual objects" (Kubovy and Schutz 2010). As discussed in Vatakis and Spence (2008), multisensory integration is modulated by both bottom-up and top-down processes (cf. Stein and Meredith 1993). Note that, in the decision-making approach suggested here, these two types of processes are captured by the likelihood function and the prior probability, respectively, suggesting the possibility of gauging their effects separately.

[3] These remarks were prompted by some of the reviewers' comments.

Acknowledgments This research is supported by grant SFB/TR31 (Project B4) from Deutsche Forschungsgemeinschaft (DFG) to H.C. and by a grant from Nowetas Foundation to both authors. We are most thankful to Marc Van Wanrooij for providing us with data on the effect of prior probability. Helpful comments from the reviewers are also gratefully acknowledged.

Appendix 1: probability of integration $P(I)$

The peripheral processing times V for the visual and A for the visual stimulus have an exponential distribution with parameters λ_V and λ_A, respectively. That is,

$$f_V(t) = \lambda_V e^{-\lambda_V t}$$
$$f_A(t) = \lambda_A e^{-\lambda_A t}$$

for $t \geq 0$, and $f_V(t) = f_A(t) \equiv 0$ for $t < 0$. The corresponding distribution functions are referred to as $F_V(t)$ and $F_A(t)$.

The visual stimulus is the target and the auditory stimulus is the nontarget. By definition,

$$\begin{aligned} P(I) &= Pr(A + \tau < V < A + \tau + \omega) \\ &= \int_0^\infty f_A(x)\{F_V(x + \tau + \omega) - F_V(x + \tau)\}dx, \end{aligned}$$

where τ denotes the SOA value and ω is the width of the integration window. Computing the integral expression requires that we distinguish between three cases for the sign of $\tau + \omega$:

(i) $\tau < \tau + \omega < 0$

$$\begin{aligned} P(I) = & \int_{-\tau-\omega}^{-\tau} \lambda_A e^{-\lambda_A x}\{1 - e^{-\lambda_V(x+\tau+\omega)}\}\mathrm{d}x \\ & + \int_{-\tau}^{\infty} \lambda_A e^{-\lambda_A x}\{e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)}\}\mathrm{d}x \\ & = \frac{\lambda_V}{\lambda_V + \lambda_A} e^{\lambda_A \tau}(-1 + e^{\lambda_A \omega}); \end{aligned}$$

(ii) $\tau < 0 < \tau + \omega$

$$\begin{aligned} P(I) = & \int_0^{-\tau} \lambda_A e^{-\lambda_A x}\left\{1 - e^{-\lambda_V(x+\tau+\omega)}\right\}\mathrm{d}x \\ & + \int_{-\tau}^{\infty} \lambda_A e^{-\lambda_A x}\left\{e^{-\lambda_V(x+\tau)} - e^{-\lambda_V(x+\tau+\omega)}\right\}\mathrm{d}x \\ & = \frac{1}{\lambda_V + \lambda_A}\left\{\lambda_A\left(1 - e^{-\lambda_V(\omega+\tau)}\right) + \lambda_V(1 - e^{\lambda_A \tau})\right\}; \end{aligned}$$

(iii) $0 < \tau < \tau + \omega$

$$P(I) = \int_0^\infty \lambda_A e^{-\lambda_A x}\{e^{-\lambda_V (x+\tau)} - e^{-\lambda_V (x+\tau+\omega)}\}\mathrm{d}x$$
$$= \frac{\lambda_A}{\lambda_V + \lambda_A}\{e^{-\lambda_V \tau} - e^{-\lambda_V (\omega+\tau)}\}.$$

Appendix 2: asymmetric-Laplace density

The asymmetric-Laplace density $\mathbf{AL}(\theta, \kappa, \sigma)$ is defined as (cf. Kotz et al. 2001, p. 137)

$$f(t|\theta,\sigma,\kappa) = \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2}\begin{cases} \exp\left(-\frac{\sqrt{2}\kappa}{\sigma}|x-\theta|\right) & \text{if } t \geq \theta, \\ \exp\left(-\frac{\sqrt{2}}{\sigma\kappa}|x-\theta|\right) & \text{if } t < \theta, \end{cases}$$

where θ is location parameter, $\kappa = 1$ implies symmetry, and σ is a scale parameter. In deriving the distributional form used in the text, the following parameter mappings have been made:

$$\sigma^2 = \frac{2}{\lambda_V \lambda_A}; \quad \kappa^2 = \lambda_V/\lambda_A; \quad \theta = 0.$$

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Anastasio TJ, Patton PE, Belkacem-Boussaid K (2000) Using Bayes‘ rule to model multisensory enhancement in the superior colliculus. Neural Comput 12:1165–1187

Battaglia PW, Jacobs RA, Aslin RN (2003) Bayesian integration of visual and auditory signals for spatial localization. J Opt Soc Am Optic Image Vis 20:1391–1397

Beierholm U, Koerding K, Shams L, Ma WJ (2008) Comparing Bayesian models for multisensory cue combination without mandatory integration. Advances in neural information processing systems, vol 20. The MIT Press, Cambridge, pp 81–88

Cohen MX (2011) It’s about time. Front Hum Neurosci 5:2. doi: 10.3389/fnhum.2011.00002

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

Colonius H, Diederich A (2010) The optimal time window of visual-auditory integration: a reaction time analysis. Front Integr Neurosci 4:11. doi:10.338/fnint.2010.00011

Colonius H, Diederich A (2006) Race model inequality: interpreting a geometric measure of the amount of violation. Psychol Rev 113(1):148–154

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

Colonius H, Diederich A (2004) Why aren’t all deep superior colliculus neurons multisensory? A Bayes’ ratio analysis. Cog Affect Behav Neurosci 4(3):344–353

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 (2011) Modeling multisensory processes in saccadic responses: time-window-of-integration model. In: Wallace MT, Murray MM (eds) Frontiers in the neural bases of multisensory processes. CRC Press, Boca Raton

Diederich A, Colonius H (2007) Why two “distractors” are better than one: modeling the effect of non-target auditory and tactile stimuli on visual saccadic reaction time. Exp Brain Res 179:43–54

Diederich A, Colonius H (2007) Modeling spatial effects in visual-tactile saccadic reaction time. Percept Psychophys 69(1):56–67

Diederich A, Colonius H (2008) 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 (2008) When a high-intensity “distractor” is better then 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 (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. The MIT Press, Cambridge, pp 395–408

Diederich A, Colonius H (1987) Intersensory facilitation in the motor component? A reaction time analysis. Psychol Res 49:23–29

Diederich A, Colonius H, Schomburg A (2008) Assessing age-related multisensory enhancement with the time-window-of-integration model. Neuropsychologia 46:2556–2562

Di Luca M, Machulla TK, Ernst MO (2009) Recalibration of multisensory simultaneity: cross-modal transfer coincides with a change in perceptual latency. J Vis 9(12):1–16

Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. J Vis 7(5):1–14

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870):429–433

Ernst MO (2005) A Bayesian view on multimodal cue integration. In: Knoblich G, Thornton I, Grosjean M, Shiffrar M (eds) Human body perception from the inside out. Oxford University Press, New York, pp 105–131

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

Giray M, Ulrich R (1993) Motor coactivation revealed by response force in divided and focused attention. J Exp Psychol Hum Percept Perform 19(6):1278–1291

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

Hillis JM, Ernst MO, Banks MS, Landy MS (2002) Combining sensory information: mandatory fusion within, but not between, senses. Science 298(5598):1627–1630

Hugenschmidt CE, Mozolic JL, Tan H, Kraft RA, Laurienti PJ (2009) Age-related increase in cross-sensory noise in resting and steady-state cerebral perfusion. Brain Topogr 21(3–4):241–251

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

Innes-Brown H, Crewther D (2009) The Impact of spatial Incongruence on an auditory-visual Illusion. PLoS ONE 4(7). doi: 10.1371/journal.pone.0006450

Keetels M, Vroomen J (2007) No effect of auditory-visual spatial disparity on temporal recalibration. Exp Brain Res 182(4):559–565

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS ONE 2(9):e943. doi:10.1371/journal.pone.0000943

Kotz S, Kozubowski TJ, Podgórski K (2001) The Laplace distribution and generalizations. A revisit with applications to communications, economics, engineering, and finance. Birkhäuser, Boston

Kubovy M, Schutz M (2010) Audio-visual objects. Rev Phil Psych 1:41–61

Laurienti PJ, Burdette JH, Maldjian JA, Wallace MT (2006) Enhanced multisensory integration in older adults. Neurobiol Aging 27:1155–1163

Lewkowicz DJ (1996) Perception of auditory-visual temporal synchrony in human infants. J Exp Psychol: Hum Percept Perform 22:1094–1106

Lewkowicz DJ (2010) Infant perception of audio-visual speech synchrony. Dev Psychol 46(1):66–77

Meredith MA (2002) On the neural basis for multisensory convergence: a brief overview. Cogn Brain Res 14:31–40

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

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a highdensity electrical mapping study. Cereb Cortex 14:452–465

Parise C, Spence C (2009) 'When birds of a feather clock together': synesthetic correspondences modulate audiovisual integration in non–synesthetes. PLoS ONE 4(5):e5664

Peiffer AM, Mozolica JL, Hugenschmidt CE, Laurienti PJ (2007) Age-related multisensory enhancement in a simple audiovisual detection task. Neuro Report 18(10):1077–1081

Poliakoff E, Shore DI, Lowe C, Spence C (2006) Visuotactile temporal order judgments in ageing. Neurosci Lett 396(3):207–211

Powers AR III, Hillock AR, Wallace MT (2009) Perceptual training narrows the temporal window of multisensory binding. J Neurosci 29(39):12265–12274

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

Roach NW, Heron J, McGraw PV (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proc Biol Sci 273:2159–2168

Roseboom W, Nishida S, Arnold DH (2009) The sliding window of audio–visual simultaneity. J Vis 9(12):4.1–8.1

Sanabria D, Soto-Faraco S, Chan JS, Spence C (2005) Intramodal perceptual grouping modulates multisensory integration: evidence from the crossmodal congruency task. Neurosci Lett 377:59–64

Sato Y, Toyoizumi T, Aihara K (2007) Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. Neural Comput 19(12):3335–3355

Setti A, Burke KE, Kenny RA, Newell FN (2011) Is inefficient multisensory processing associated with falls in older people. Exp Brain Res 209(3):375–384

Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. Nature 408(6814):788

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14(1):147–152

Shams L, Ma WJ, Beierholm U (2005) Sound-induced flash illusion as an optimal percept. Neuroreport 16:1923–1927

Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. Cognit Brain Res 14:139–146

Spence C (2007) Audiovisual multisensory integration. Acoust Sci Technol 28:61–70

Stein BE, Meredith MA (1993) The merging of the senses. The MIT Press, Cambridge

Van Wanrooij MM, Bell AH, Munoz DP, Van Opstal AJ (2009) The effect of spatial-temporal audiovisual disparities on saccades in a complex scene. Exp Brain Res 198:425–437

Van Wanrooij MM, Bremen P, Van Opstal AJ (2010) Acquired prior knowledge modulates audiovisual integration. Eur J Neurosci 31:1763–1771

Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. Attent Percept Psychophys 72:871–884

Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. Brain Res Cogn Brain Res 22:32–35

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. Exp Brain Res 158(2):252–258

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

Wozny D, Beierholm U, Shams L (2010) Probability matching as a computational strategy used in perception. PLoS Comput Biol 6(8):e1000871. doi:10.1371/journal.pcbi.1000871