Psychon Bull Rev (2012) 19:820–846
DOI 10.3758/s13423-012-0278-y

THEORETICAL REVIEW

On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model

Miguel A. García-Pérez · Rocío Alcalá-Quintana

Published online: 25 July 2012

Abstract Research on the perception of temporal order uses either *temporal-order judgment* (TOJ) tasks or *synchrony judgment* (SJ) tasks, in both of which two stimuli are presented with some temporal delay and observers must judge the order of presentation. Results generally differ across tasks, raising concerns about whether they measure the same processes. We present a model including sensory and decisional parameters that places these tasks in a common framework that allows studying their implications on observed performance. TOJ tasks imply specific decisional components that explain the discrepancy of results obtained with TOJ and SJ tasks. The model is also tested against published data on audiovisual temporal-order judgments, and the fit is satisfactory, although model parameters are more accurately estimated with SJ tasks. Measures of *latent point of subjective simultaneity* and *latent sensitivity* are defined that are invariant across tasks by isolating the sensory parameters governing observed performance, whereas decisional parameters vary across tasks and account for observed differences across them. Our analyses concur with other evidence advising against the use of TOJ tasks in research on perception of temporal order.

Keywords Synchrony judgment · Temporal-order judgment · Point of subjective simultaneity · Audiovisual events · Experimental methods · Stimulus type

Electronic supplementary material The online version of this article (doi:10.3758/s13423-012-0278-y) contains supplementary material, which is available to authorized users.

M. A. García-Pérez (✉) · R. Alcalá-Quintana
Departamento de Metodología, Facultad de Psicología,
Universidad Complutense,
Campus de Somosaguas,
28223 Madrid, Spain
e-mail: miguel@psi.ucm.es

Research on perception of temporal order investigates observers' ability to elucidate whether two sensory events occurred simultaneously or in what order they occurred. This ability is measured as a function of the temporal delay (or stimulus onset asynchrony; SOA) between the two events, and this strategy is used with events of the same or of different sensory modalities. Here, we will refer only to the audiovisual case for simplicity and without loss of generality.

In studies of audiovisual synchrony or temporal-order judgments, the onset of the visual signal is usually regarded as the reference, and the independent variable is *auditory delay*, defined as positive when the auditory signal follows the visual signal and as negative when the auditory signal precedes the visual signal. Those studies use one of two tasks. In the *synchrony judgment* (SJ) task, the observer indicates whether the stimuli were presented simultaneously or successively, yielding *synchronous* (S) or *asynchronous* (A) responses in the SJ2 version of the task, or whether the auditory stimulus was presented before, after, or at the same time as the visual stimulus, yielding *audio-first* (AF), *video-first* (VF), or *synchronous* (S) responses in the SJ3 version. In the *temporal-order judgment* (TOJ) task, the observer must judge which stimulus was presented first (yielding AF or VF responses). The data are typically represented as psychometric functions reflecting the proportion of responses of each type as a function of auditory delay. Examples of empirical psychometric functions under each type of task are given in Figs. 5 and 6 below.

Two quantitative indices are typically obtained from SJ data. One is the *synchrony range*, which is the width of the interval of temporal delays for which S responses prevail. The other is the *point of subjective simultaneity* (PSS), which is the auditory delay at the midpoint of the synchrony interval. TOJ tasks also yield a PSS (the auditory delay at

which AF and VF responses are equally frequent) and a sensitivity measure (the slope of the psychometric function at the PSS). Van Eijk, Kohlrausch, Juola, and van de Par (2008) tabulated results from studies aimed at estimating the PSS with SJ and TOJ tasks and showed that the two tasks generally provide discrepant estimates. Van Eijk et al. (2008) also carried out a within-subjects study involving the three tasks (SJ2, SJ3, and TOJ), whose results confirmed that PSS estimates from SJ2 and SJ3 tasks were highly correlated, whereas PSS estimates from TOJ tasks were uncorrelated with those from SJ2 or SJ3 tasks. Their results also revealed that PSS estimates from SJ2 and SJ3 tasks did not differ significantly, whereas PSS estimates from TOJ tasks were significantly lower than those from SJ2 or SJ3 tasks. The reason for the discrepant PSS estimates in SJ and TOJ tasks is unclear, but these discrepancies have prompted the view that SJ and TOJ tasks measure distinct processes (for a review, see Spence & Parise, 2010). The consensus is that SJ and TOJ tasks imply different response biases (e.g., Nicholls, Lew, Loetscher, & Yates, 2011; Spence & Parise, 2010; Vatakis, Navarra, Soto-Faraco, & Spence, 2008; Vroomen & Keetels, 2010; Yates & Nicholls, 2011), but their nature has never been specified and how they affect performance has never been modeled. The issue of whether SJ and TOJ tasks involve different processes lends itself to investigation with recourse to models of timing judgments, in which sensory and decisional aspects (including biases) are explicitly represented.

Most models of timing judgments fall within the class of independent-channels models described by Sternberg and Knoll (1973).[1] In these models, signals from the two stimuli arrive at a central mechanism with randomly distributed *arrival latencies*. At this mechanism, the judgment of temporal order or synchrony is determined by a ternary *decision rule* applied to the *arrival-time difference* between the two signals. Sternberg and Knoll considered six variants of the decision rule (which accommodate the class of attention-switching models of Kristofferson & Allan, 1973) and analyzed the formal properties of these models mostly as regards reaction times under TOJ tasks. Sternberg and Knoll also derived general properties of independent-channels models.

Research on timing judgments has relied on models to various extents. Some studies fitted specific models to data from a two-alternative forced-choice (2AFC) variant of the TOJ task (e.g., Allan & Kristofferson, 1974) or to data from SJ or TOJ tasks (e.g., Allan, 1975; Jaśkowski, 1991a, 1991b, 1993). In later studies, models were mentioned but subsequently obviated, and arbitrary functions were instead fitted to data with the goal of obtaining measures of PSS and sensitivity from the fitted functions (e.g., Heath, 1984; Stelmach & Herdman, 1991). The most recent research disregards models entirely and merely fits arbitrary functions to SJ and TOJ data, with no connection to explicit models (e.g., Donohue, Woldorff, & Mitroff, 2010; Fujisaki & Nishida, 2009; Harrar & Harris, 2008; Nicholls et al., 2011; Shore, Spry, & Spence, 2002; Spence, Baddeley, Zampini, James, & Shore, 2003; Stone et al., 2001; van Eijk et al., 2008, 2010; Vatakis et al., 2008; Yamamoto & Kitazawa, 2001; Yates & Nicholls, 2011; Zampini, Guest, Shore, & Spence, 2005; Zampini, Shore, & Spence, 2003). An exception is the study by Schneider and Bavelier (2003), whose goal was distinguishing mechanisms (response biases or criterion changes) that might masquerade as prior entry effects. They thus looked into models from this perspective only and, hence, did not assess how model parameters (as opposed to derived measures such as the PSS) differed between SJ and TOJ tasks.

An important consequence of the widespread practice of fitting arbitrary functions to data is the resultant multiplicity of uninterpretable parameters. Consider the study of van Eijk et al. (2008), in which the same stimuli and SOAs were used under the three tasks. Van Eijk et al. (2008) fitted two independent 3-parameter functions to SJ2 data, four independent 3- or 4-parameter functions to SJ3 data, and a further independent 4-parameter function to TOJ data. Thus, 24 parameters were involved, which reflected only the location, slope, and asymptotes of the curves that best described the path of each (sub)set of data. Since stimuli and conditions were the same across tasks, sensory aspects (i.e., distributions of arrival latencies) are likely to have been invariant across tasks, with differences across them only in criteria or biases. An analogous situation arises in studies in which a given task is used under different attentional conditions (e.g., when cuing or other manipulations are used in studies on prior entry). In these cases, criteria or biases are likely to be the same across manipulations, which should presumably alter only sensory parameters. The strategy of fitting arbitrary functions to data (which usually renders different PSSs, as discussed above) cannot shed light on these issues, because there is no link between the parameters of the arbitrary curves fitted to the data and the sensory and decisional parameters governing judgments.

The work described in this article looked into these issues, and the plan of the article is as follows. First, an explicit model is presented that places performance in all tasks in a common framework also incorporating response errors. The model is an expanded version of Sternberg and Knoll's (1973) independent-channels model 3. A formal analysis reveals implications for performance measures in SJ and TOJ tasks and also shows that the model can accommodate a number of quantitative and qualitative differences

[1] Consideration of a distinctively different class of models assuming that perception of simultaneity and perception of temporal order follow separate and independent processing streams will be deferred to the Discussion section.

between SJ and TOJ data. The model is then used to test hypotheses about differences in sensory or decisional parameters across tasks. To this end, the model is fitted to the data of van Eijk et al. (2008), both under no constraint (i.e., SJ and TOJ data are fitted separately, which may render different parameter estimates across tasks) and under the constraint that sensory parameters are common, whereas decisional parameters differ across tasks. When the model was fitted separately, estimated sensory parameters turned out to be similar in all tasks, whereas estimated decisional parameters differed across tasks, providing justification for the constrained fit. We discuss the interpretation of model parameters and extract conventional performance measures from the data, which confirm the discrepancies that van Eijk et al. (2008) reported using a different method of analysis. Finally, we discuss several extensions to the model, as well as alternative and task-independent measures of timing judgments.

An integrated model for SJ2, SJ3, and TOJ tasks

The model to be presented next falls within the class of independent-channels models of Sternberg and Knoll (1973) and is similar to the triggered-moment model of Schneider and Bavelier (2003), although our model differs in some respects. A preliminary version of the model for application to SJ3 tasks has been described (García-Pérez & Alcalá-Quintana, 2012a), which showed that the model can account for aspects of SJ3 data that were previously regarded as evidence against independent-channels models. The keystone of the model is the distinction between unobservable judgments and observed responses, and the model is extended here to SJ2 and TOJ tasks, as suggested by Schneider and Bavelier. The extension is built on the surmise that an observer's judgment of whether two stimuli were presented simultaneously or in a given order precedes and is independent of the response requested from the observer. Thus, the model assumes that, on each trial, the observer strives to collect sensory information so as to judge whether the visual stimulus was first or the auditory stimulus was first; if the observer cannot tell, presentation is judged to be simultaneous. Thus, observers face each trial *as if* they were to choose among these three options and, subsequently, translate this judgment into a response according to the options given in each task: AF, S, or VF in SJ3 tasks, only S or A in SJ2 tasks, and only AF or VF in TOJ tasks. In giving these responses, observers may make *finger errors* and, thus, misreport their judgments. The model will be presented in two steps: first, the model for unobservable judgments and, then, the model for observed responses, including potential misreports due to finger errors.

Let the *arrival latencies* T_v and T_a of visual and auditory signals be random variables with respective densities g_v and g_a given by the shifted exponential distributions

$$g_i(t) = \lambda_i \exp[-\lambda_i(t - (\Delta t_i + \tau_i))], \quad t \geq \Delta t_i + \tau_i,\ i \in \{\mathrm{v},\ \mathrm{a}\}, \tag{1}$$

where Δt_i (in milliseconds) is the actual onset of signal i, λ_i is the rate parameter (in milliseconds^{-1}), and τ_i (in milliseconds) is a further processing delay. Exponential distributions are commonly assumed to describe arrival latencies or peripheral processing times (see, e.g., Colonius & Diederich, 2011; Diederich & Colonius, 2011; Heath, 1984), and their parameters are readily interpretable: The mean of a variable distributed as in Eq. 1 is $1/\lambda_i + \tau_i$, whereas the standard deviation is $1/\lambda_i$. Thus, parameters λ_i and τ_i describe arrival latencies whose mean and standard deviation may vary independently.

Without loss of generality, we set the arbitrary origin of time at the physical onset of the visual stimulus so that $\Delta t_v = 0$ and, hence, $\Delta t \equiv \Delta t_a$ is the auditory delay that is manipulated, where $\Delta t < 0$ reflects that the auditory signal precedes the visual signal, and $\Delta t > 0$ reflects that the auditory signal follows the visual signal. Figure 1a shows an example. On a given trial, arrival latencies are realizations of these distributions, and the observer's judgment arises from a decision rule operating on the arrival-time difference $D = T_a - T_v$, which has a bilateral exponential distribution given

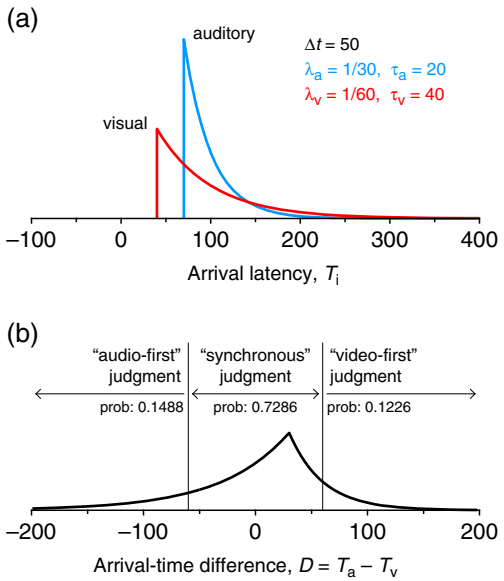

Fig. 1 Model of timing judgments. **a** Exponential distributions for the arrival latency of a visual stimulus (red curve) presented at time 0 and an auditory stimulus (blue curve) presented at time Δt = 50 ms. Parameters as indicated in the inset. **b** Bilateral exponential distribution of arrival-time differences and cutpoints on the decision space (vertical lines, at $D = \pm\delta$ with δ = 60), determining the probability of each judgment

by

$$f(d;\Delta t)=\begin{cases}\frac{\lambda_a\lambda_v}{\lambda_a+\lambda_v}\exp[\lambda_v(d-\Delta t-\tau)] & \text{if } d\le\Delta t+\tau\\ \frac{\lambda_a\lambda_v}{\lambda_a+\lambda_v}\exp[-\lambda_a(d-\Delta t-\tau)] & \text{if } d>\Delta t+\tau\end{cases}, \tag{2}$$

where $\tau=\tau_a-\tau_v$ is a *processing advantage* such that $\tau<0$ indicates faster auditory processing and $\tau>0$ indicates faster visual processing. Figure 1b shows the distribution for the case in Fig. 1a.

The decision process operates under the trichotomy defined by the two vertical lines in Fig. 1b. Thus, an AF judgment arises when D is sufficiently large and negative ($D<-\delta$), a VF judgment arises when D is sufficiently large and positive ($D>\delta$), and an S judgment occurs when the arrival-time difference is sufficiently small ($-\delta\le D\le\delta$). Here, δ is a *resolution* parameter determining the observer's ability to discriminate small differences in arrival latency.[2] Under the SJ2 task, observers using the same decision space arrive at A judgments whenever AF or VF judgments occur. Finally, TOJ tasks only allow VF or AF responses and, hence, force observers to guess between AF and VF responses when they make S judgments. Let ξ be the probability of a VF response in these cases, so that $1-\xi$ is the probability of an AF response. Then, in TOJ tasks, VF responses occur when $D>\delta$ (as a result of VF judgments) and with probability ξ when $-\delta\le D\le\delta$ (as a result of guesses), whereas AF responses occur when $D<-\delta$ and with probability $1-\xi$ when $-\delta\le D\le\delta$. We will refer to ξ as a *response bias* parameter because it may produce an imbalance of guessing outcomes due to a bias toward one of the response options.

Even if parameters do not vary across tasks, this model predicts differences across tasks as to how the probability of judgments of synchrony and temporal order vary as a function of Δt. To obtain these predictions, first note that the cumulative distribution for D is

$$\begin{aligned}F(d;\Delta t)&=\int_{-\infty}^{d}f(z;\Delta t)\,dz\\&=\begin{cases}\frac{\lambda_a}{\lambda_a+\lambda_v}\exp[\lambda_v(d-\Delta t-\tau)] & \text{if } d\le\Delta t+\tau\\ 1-\frac{\lambda_v}{\lambda_a+\lambda_v}\exp[-\lambda_a(d-\Delta t-\tau)] & \text{if } d>\Delta t+\tau\end{cases},\end{aligned} \tag{3}$$

where f is given by Eq. 2. Let $\Psi_{X\text{-}Y}$ be the probability under task X (for X $\in$ {SJ2, SJ3, TOJ}) of judgment Y (for Y $\in$ {VF, AF, S, A}). Then,

$$\Psi_{SJ3\text{-}AF}(\Delta t)=\int_{-\infty}^{-\delta}f(z;\Delta t)\,dz=F(-\delta;\Delta t) \tag{4a}$$

$$\Psi_{SJ3\text{-}S}(\Delta t)=\int_{-\delta}^{\delta}f(z;\Delta t)\,dz=F(\delta;\Delta t)-F(-\delta;\Delta t) \tag{4b}$$

$$\Psi_{SJ3\text{-}VF}(\Delta t)=\int_{\delta}^{\infty}f(z;\Delta t)\,dz=1-F(\delta;\Delta t) \tag{4c}$$

in the SJ3 task,

$$\Psi_{SJ2\text{-}S}(\Delta t)=\int_{-\delta}^{\delta}f(z;\Delta t)\,dz=\Psi_{SJ3\text{-}S}(\Delta t) \tag{5a}$$

$$\begin{aligned}\Psi_{SJ2\text{-}A}(\Delta t)&=\int_{-\infty}^{-\delta}f(z;\Delta t)\,dz+\int_{\delta}^{\infty}f(z;\Delta t)\,dz\\&=\Psi_{SJ3\text{-}AF}(\Delta t)+\Psi_{SJ3\text{-}VF}(\Delta t)\end{aligned} \tag{5b}$$

in the SJ2 task, and

$$\begin{aligned}\Psi_{TOJ\text{-}AF}(\Delta t)&=\int_{-\infty}^{-\delta}f(z;\Delta t)\,dz+(1-\xi)\int_{-\delta}^{\delta}f(z;\Delta t)\,dz\\&=\Psi_{SJ3\text{-}AF}(\Delta t)+(1-\xi)\Psi_{SJ3\text{-}S}(\Delta t)\end{aligned} \tag{6a}$$

$$\begin{aligned}\Psi_{TOJ\text{-}VF}(\Delta t)&=\int_{\delta}^{\infty}f(z;\Delta t)\,dz+\xi\int_{-\delta}^{\delta}f(z;\Delta t)\,dz\\&=\Psi_{SJ3\text{-}VF}(\Delta t)+\xi\,\Psi_{SJ3\text{-}S}(\Delta t)\end{aligned} \tag{6b}$$

in the TOJ task. Psychometric functions under SJ2 and TOJ tasks are thus transformations of the psychometric functions for the SJ3 task, although they describe different shapes. Figure 2 shows these functions with the parameters in Fig. 1 and for $\xi\in\{0, .2, .5, .8, 1\}$. Appendix 1 discusses how model parameters determine the indices of performance extracted when these functions are fitted to empirical data (synchrony boundaries, synchrony ranges, and PSSs).

Model parameters must vary across individuals, but they may also vary with stimulus conditions (e.g., cuing or the availability of anticipatory sensory information, which might alter λ_v, λ_a, τ_v, or τ_a), with learning, fatigue, or task requirements (which may alter the resolution parameter δ), or with strategic decisions or expectations (which may alter the response bias parameter ξ in TOJ tasks). Figure 3 illustrates two extreme cases involving changes in δ. In the top panel, resolution is high (i.e., δ is small), as it would be for an observer with high ability to resolve small differences in arrival latency. Due to this ability, S judgments in the SJ3

[2] The reason for a single parameter to define a symmetric central zone in the decision space under independent-channel models is that an asymmetric zone requires two independent boundary parameters whose effects are not experimentally distinguishable from those of parameter τ. For details, see Appendix A.1 in Schneider and Bavelier (2003) and the extensive discussion in Yarrow, Jahn, Durant, and Arnold (2011).

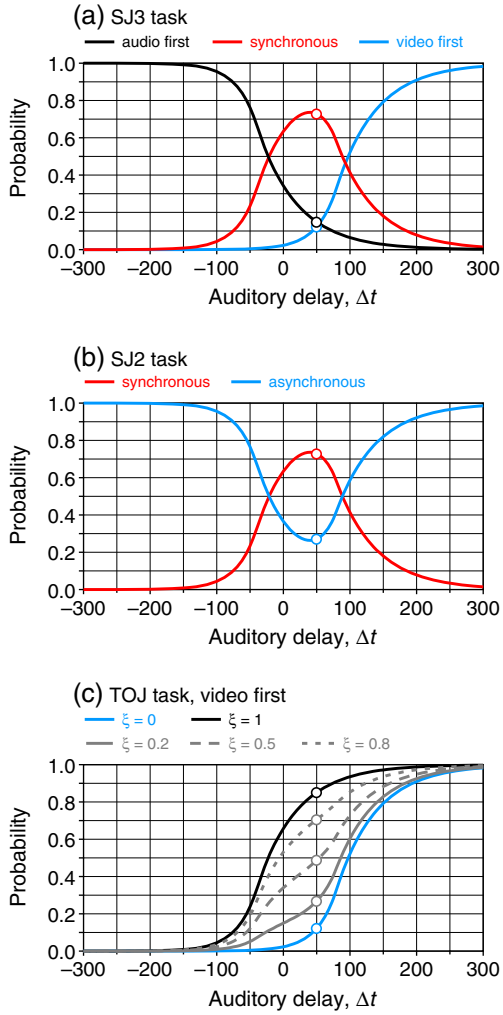

Fig. 2 Probability of each judgment in each task as a function of auditory delay. **a** SJ3 task. **b** SJ2 task. **c** TOJ task. Parameters of the underlying exponential are the same in all cases and equal to those given in Fig. 1. The various curves in panel c indicate how the curve describing the probability of VF responses (judgments plus guesses) in the TOJ task varies with response bias (parameter ξ). The circle on each curve in each panel reflects the probability of each judgment under each task when Δt = 50 ms. Thus, in panel a, the probability of VF, S, and AF judgments are .1266, .7286, and .1488, respectively (as shown in Fig. 1b); in panel b, the probability of an S judgment is still .7286, and the probability of an A judgment is .1226 + .1488 = .2714; in panel c, the probability of a VF response (a mixture of true judgments and guesses) varies with bias: If ξ = .5 (i.e., an unbiased observer), the probability is .1226 + .5 × .7286 = .4869; if ξ = .8 (i.e., an observer biased toward VF responses), the probability is instead .1226 + .8 × .7286 = .7055; if ξ = .2 (i.e., an observer biased toward AF responses), the probability is .1226 + .2 × .7286 = .2683; finally, only if ξ = 0 (i.e., for an observer fully biased toward AF responses) will the probability of a VF response in the TOJ task equal the probability of a VF judgment in the SJ3 task, whereas if ξ = 1 (i.e., an observer fully biased toward VF responses), the probability of a VF response in the TOJ task equals one minus the probability of an AF judgment in the SJ3 task

task are infrequent, and $\Psi_{SJ3\text{-}S}$ (red curve) does not cross $\Psi_{SJ3\text{-}VF}$ (blue curve). This characteristic was reported by Stelmach and Herdman (1991; see their Fig. 5) and reflects what van Eijk et al. (2008) described as "synchronous response proportions [that] were always below either the audio-first or the video-first curve" (p. 962). At the same time, and because of the scarcity of S judgments, response bias in the TOJ task has a very limited effect, and variations in response bias produce functions $\Psi_{TOJ\text{-}VF}$ (gray curves) that differ only minimally from their sister function $\Psi_{SJ3\text{-}VF}$ (blue curve).

The bottom panel of Fig. 3 shows a case differing only in that resolution is now poor (i.e., δ is large). S judgments now prevail in the SJ3 task (red curve), producing a pattern akin to what van Eijk et al. (2010) described as "synchronous responses [that] were relatively constant over the range of relative delays" (p. 2231). The prevalence of S judgments makes response bias affect the shape of $\Psi_{TOJ\text{-}VF}$ (gray curves), which may thus differ greatly from that of $\Psi_{SJ3\text{-}VF}$ (blue curve).

The extreme cases in Fig. 3 help describe how the functions change for intermediate values of δ. As δ increases, $\Psi_{SJ3\text{-}S}$ progressively broadens and rises, in a transition from the red curve in the top panel of Fig. 3 to the red curve in the bottom panel of Fig. 3; analogously, $\Psi_{SJ3\text{-}VF}$ progressively shifts to the right in a transition from the blue curve in the top panel of Fig. 3 to the blue curve in the bottom panel of Fig. 3; and, finally, the effect of the response bias parameter ξ increases and makes $\Psi_{TOJ\text{-}VF}$ (gray curves in Fig. 3) describe a transition from the shapes in the top panel (where

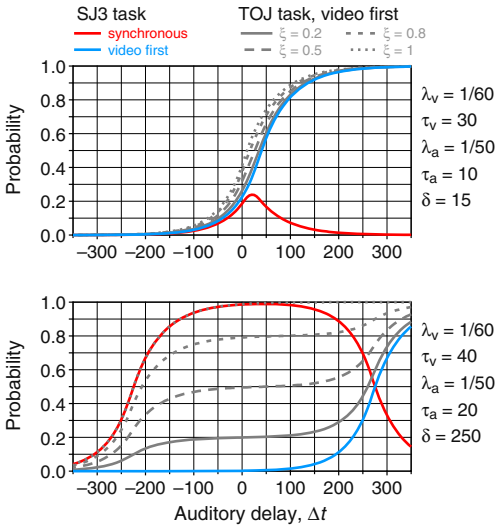

Fig. 3 Model curves for the SJ3 and TOJ tasks under extreme cases regarding the resolution parameter δ. For small δ (δ = 15; top panel), S judgments in the SJ3 task (red curve) are relatively infrequent, and hence, the curve for S judgments does not cross the curve for VF judgments (blue curve). At the same time, variations in response bias (parameter ξ) produce patterns of VF responses in the TOJ task (gray curves) that barely differ from VF judgments in the SJ3 task (blue curve). For larger δ (δ = 250; bottom panel), S judgments prevail in the SJ3 task (red curve) and variations in response bias (parameter ξ) produce patterns of VF responses in the TOJ task (gray curves) that differ greatly from VF judgments in the SJ3 task (blue curve)

curves for different ξs are similar and tightly packed together) to those in the bottom panel (where curves for different ξs spread out horizontally and differ in the height of their intermediate plateau).

Up to now, model equations express the probabilities of unobservable judgments (plus guesses in TOJ tasks) as a function of auditory delay. These equations express probabilities of observed responses only if responses faithfully reflect judgments. However, errors on pressing the response keys decouple observed responses and judgments, and it has been shown that the presumed failure of independent-channels models can be explained as a result of these errors (García-Pérez & Alcalá-Quintana, 2012a).[3] Thus, let ε_{AF}, ε_{S}, and ε_{VF} be the probabilities that the observer misreports AF, S, and VF judgments, respectively. In SJ3 tasks, misreporting any given judgment can take two forms (i.e., an observer who has made an S judgment may unintentionally press the response key for an AF or a VF judgment). Let $\kappa_{X\text{-}Y}$ be the bias toward misreporting an X judgment as a Y response so that $\kappa_{X\text{-}Z} = 1 - \kappa_{X\text{-}Y}$ is the bias toward misreporting an X judgment as a Z response. These biases rarely reflect observers' decisions; on the contrary, they are often caused by the layout of the response interface, which may make some observers more prone to making errors upon reporting specific judgments and, in such cases, more prone to misreporting in a certain form. Only three such bias parameters exist—say, $\kappa_{AF\text{-}S}$, $\kappa_{S\text{-}AF}$, and $\kappa_{VF\text{-}AF}$—because $\kappa_{AF\text{-}VF} = 1 - \kappa_{AF\text{-}S}$, $\kappa_{S\text{-}VF} = 1 - \kappa_{S\text{-}AF}$, and $\kappa_{VF\text{-}S} = 1 - \kappa_{VF\text{-}AF}$. The tree diagrams in Fig. 4 indicate the possible sequences of events mapping judgments onto responses when errors may occur. These diagrams also help demonstrate that the model incorporating response errors in SJ3 tasks is

$$\begin{aligned}\Psi^{*}_{SJ3\text{-}AF}(\Delta t) &= (1-\varepsilon_{SJ3\text{-}AF})\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+\varepsilon_{SJ3\text{-}S}\kappa_{S\text{-}AF}\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+\varepsilon_{SJ3\text{-}VF}\kappa_{VF\text{-}AF}\Psi_{SJ3\text{-}VF}(\Delta t)\end{aligned} \tag{7a}$$

$$\begin{aligned}\Psi^{*}_{SJ3\text{-}S}(\Delta t) &= \varepsilon_{SJ3\text{-}AF}\kappa_{AF\text{-}S}\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+(1-\varepsilon_{SJ3\text{-}S})\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+\varepsilon_{SJ3\text{-}VF}(1-\kappa_{VF\text{-}AF})\Psi_{SJ3\text{-}VF}(\Delta t)\end{aligned} \tag{7b}$$

$$\begin{aligned}\Psi^{*}_{SJ3\text{-}VF}(\Delta t) &= \varepsilon_{SJ3\text{-}AF}(1-\kappa_{AF\text{-}S})\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+\varepsilon_{SJ3\text{-}S}(1-\kappa_{S\text{-}AF})\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+(1-\varepsilon_{SJ3\text{-}VF})\Psi_{SJ3\text{-}VF}(\Delta t),\end{aligned} \tag{7c}$$

where $\Psi_{SJ3\text{-}AF}$, $\Psi_{SJ3\text{-}S}$, and $\Psi_{SJ3\text{-}VF}$ on the right-hand sides are given by Eqs. 4a, 4b and 4c. Similarly, the extended model for SJ2 tasks is

$$\begin{aligned}\Psi^{*}_{SJ2\text{-}S}(\Delta t) &= \varepsilon_{SJ2\text{-}AF}\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+(1-\varepsilon_{SJ2\text{-}S})\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+\varepsilon_{SJ2\text{-}VF}\Psi_{SJ3\text{-}VF}(\Delta t)\end{aligned} \tag{8a}$$

$$\begin{aligned}\Psi^{*}_{SJ2\text{-}A}(\Delta t) &= (1-\varepsilon_{SJ2\text{-}AF})\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+\varepsilon_{SJ2\text{-}S}\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+(1-\varepsilon_{SJ2\text{-}VF})\Psi_{SJ3\text{-}VF}(\Delta t),\end{aligned} \tag{8b}$$

and the extended model for TOJ tasks is

$$\begin{aligned}\Psi^{*}_{TOJ\text{-}AF}(\Delta t) &= (1-\varepsilon_{TOJ\text{-}AF})\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+(1-\xi)\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+\varepsilon_{TOJ\text{-}VF}\Psi_{SJ3\text{-}VF}(\Delta t)\end{aligned} \tag{9a}$$

$$\begin{aligned}\Psi^{*}_{TOJ\text{-}VF}(\Delta t) &= \varepsilon_{TOJ\text{-}AF}\Psi_{SJ3\text{-}AF}(\Delta t)\\ &\quad+\xi\,\Psi_{SJ3\text{-}S}(\Delta t)\\ &\quad+(1-\varepsilon_{TOJ\text{-}VF})\Psi_{SJ3\text{-}VF}(\Delta t),\end{aligned} \tag{9b}$$

where additional subscripts are used for ε_{AF}, ε_{S}, and ε_{VF} because these parameters may vary across tasks (e.g., misreports may be more frequent in ternary SJ3 tasks than in binary SJ2 or TOJ tasks).

Testing for differences across tasks

To test for differences (or lack thereof) in model parameters across tasks, the model was fitted to data reported by van Eijk et al. (2008), because their experiment used the three tasks under the same stimulus conditions and with the same observers. The experiment involved two types of stimuli. In the *flash–click* configuration, the visual component was a white disk flashed for 12 ms against a dark background, and the auditory component was a 12-ms white-noise burst. In the *bouncing ball* configuration, the visual component was an animation in which a white disk moved down toward a horizontal bar and bounced back up upon impact with the bar, whereas the auditory component was a sinusoid with a sharp onset and an exponential decay. With each configuration, auditory delays ranged between −350 and 350 ms in steps of

[3] There is an additional source of response contamination that comes from what is generally referred to as *lapses*. These refer to cases in which observers blink or have lapses of attention that make them miss the trial. The ensuing failure to collect sensory information on those trials precludes judgments, and the observer's response is a pure guess. Although this eventuality can be included in the model (see García-Pérez & Alcalá-Quintana, 2012a), we will leave lapses aside, because experiments can be designed so as to allow observers to abort trials on which they missed the stimulus for one or another reason. This commendable practice has been used, for instance, by Yarrow et al. (2011).

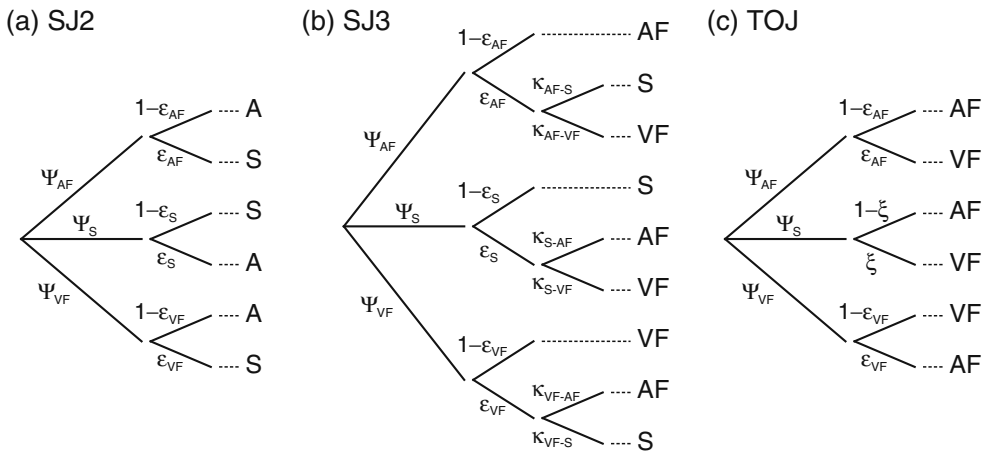

Fig. 4 Tree diagrams describing the sequences of events mapping unobservable judgments onto observed responses in the SJ2 (**a**), SJ3 (**b**), and TOJ (**c**) tasks. The starting point at the left of each diagram is the unobservable judgments, which occur with probabilities given by Eqs. 4a, 4b and 4c. Subsequently, misreports occur with probabilities given by the ε parameters. In the SJ2 task (panel a), misreports result in a response opposite to that which reflects the judgment. In the SJ3 task (panel b), misreports also render a response different from the judgment, but, since there are two of those, which one it is varies with the response bias parameters κ. In the TOJ task (panel c), misreports also affect AF or VF judgments as in the SJ2 task, whereas S judgments render an AF or a VF response at random and according to the response bias parameter ξ.

50 ms, and 60 trials were administered at each delay. The experiment was repeated under SJ2, SJ3, and TOJ tasks. Eleven observers took part in the experiment with each pairing of configuration and task, and an additional observer participated in the experiment with the bouncing ball stimulus.

We used maximum-likelihood methods to estimate model parameters for each observer and stimulus configuration, under two approaches. First, we fitted the model separately to data from each task, which may result in different parameter estimates across tasks. As will be shown below, estimated sensory parameters (λ_a, λ_v, and τ) did not differ significantly or meaningfully across tasks under this separate fit. Thus, we subsequently fitted the model jointly to data from all tasks under the constraint that λ_a, λ_v, and τ have the same values on all tasks, whereas all other parameters might differ across tasks. Pearson's chi-square statistic X^2 and the log-likelihood ratio statistic G^2 were computed as indices of fit in each case, and a model comparison approach was also used to determine whether the joint fit or the separate fit accounts better for each observer's data.

Separate fit

In the separate approach, for each observer and stimulus configuration, SJ2 model parameters were sought by maximizing the likelihood function

$$L_{\mathrm{SJ2}}(\boldsymbol{R}_{\mathrm{SJ2}};\boldsymbol{\theta}_{\mathrm{SJ2}}) = \prod_{i=1}^{N_{\mathrm{SJ2}}}\left[\Psi^*_{\mathrm{SJ2\text{-}AS}}(\Delta t_i)\right]^{A_i}\left[\Psi^*_{\mathrm{SJ2\text{-}S}}(\Delta t_i)\right]^{S_i}, \quad (10)$$

where $\boldsymbol{R}_{\mathrm{SJ2}}$ is the set of empirical responses, $\theta_{\mathrm{SJ2}} = (\lambda_a, \lambda_v, \tau, \delta, \varepsilon_{\mathrm{SJ2\text{-}AF}}, \varepsilon_{\mathrm{SJ2\text{-}S}}, \varepsilon_{\mathrm{SJ2\text{-}VF}})$ is the vector of parameters, $\{\Delta t_1, \Delta t_2, \ldots, \Delta t_{N_{\mathrm{SJ2}}}\}$ is the set of N_{SJ2} auditory delays at which responses were collected, and A_i and S_i are the counts of A and S responses given at delay Δt_i. Similarly, SJ3 model parameters were sought by maximizing

$$L_{\mathrm{SJ3}}(\boldsymbol{R}_{\mathrm{SJ3}};\boldsymbol{\theta}_{\mathrm{SJ3}}) = \prod_{i=1}^{N_{\mathrm{SJ3}}}\left[\Psi^*_{\mathrm{SJ3\text{-}AF}}(\Delta t_i)\right]^{A_i}\left[\Psi^*_{\mathrm{SJ3\text{-}S}}(\Delta t_i)\right]^{S_i}\left[\Psi^*_{\mathrm{SJ3\text{-}VF}}(\Delta t_i)\right]^{V_i}, \quad (11)$$

where $\boldsymbol{R}_{\mathrm{SJ3}}$ is the set of empirical responses, $\theta_{\mathrm{SJ3}} = (\lambda_a, \lambda_v, \tau, \delta, \varepsilon_{\mathrm{SJ3\text{-}AF}}, \varepsilon_{\mathrm{SJ3\text{-}S}}, \varepsilon_{\mathrm{SJ3\text{-}VF}}, \kappa_{\mathrm{AF\text{-}S}}, \kappa_{\mathrm{S\text{-}AF}}, \kappa_{\mathrm{VF\text{-}AF}})$ is the vector of parameters, $\{\Delta t_1, \Delta t_2, \ldots, \Delta t_{N_{\mathrm{SJ3}}}\}$ is the set of N_{SJ3} auditory delays at which responses were collected, and A_i, S_i, and V_i are the counts of AF, S, and VF responses at delay Δt_i. Finally, TOJ model parameters were sought by maximizing

$$L_{\mathrm{TOJ}}(\boldsymbol{R}_{\mathrm{TOJ}};\boldsymbol{\theta}_{\mathrm{TOJ}}) = \prod_{i=1}^{N_{\mathrm{TOJ}}}\left[\Psi^*_{\mathrm{TOJ\text{-}AF}}(\Delta t_i)\right]^{A_i}\left[\Psi^*_{\mathrm{TOJ\text{-}VF}}(\Delta t_i)\right]^{V_i}, \quad (12)$$

where $\boldsymbol{R}_{\mathrm{TOJ}}$ is the set of empirical responses, $\theta_{\mathrm{TOJ}} = (\lambda_a, \lambda_v, \tau, \delta, \varepsilon_{\mathrm{TOJ\text{-}AF}}, \xi, \varepsilon_{\mathrm{TOJ\text{-}VF}})$ is the vector of parameters, $\{\Delta t_1, \Delta t_2, \ldots, \Delta t_{N_{\mathrm{TOJ}}}\}$ is the set of N_{TOJ} auditory delays at which responses were collected, and A_i and V_i are the counts of AF and VF responses at delay Δt_i.

Likelihood functions were maximized using NAG subroutine E04JYF (Numerical Algorithms Group, 1999), which

uses a quasi-Newton algorithm for constrained optimization. As applicable according to the function to be maximized, the parameter space spanned the range [1/200, 1/5] for λ_a and λ_v, [0, 300] for δ, [−150, 150] for τ, [0, .8] for ε_{AF}, ε_S, and ε_{VF}, and [0, 1] for $\kappa_{AF\text{-}S}$, $\kappa_{S\text{-}AF}$, $\kappa_{VF\text{-}AF}$, and ξ. Two or three evenly spaced initial values were defined for each parameter. These initial values were factorially combined, and the maximization routine ran for each of the resultant starting points, yielding in each case a vector of estimates and a divergence index. On completion, we took the vector of estimates for which divergence was lowest, although the returned solution rarely differed meaningfully (if at all) across starting points (see also Alcalá-Quintana & García-Pérez, 2012).

The process just described estimates parameters for the *full* model for each task—that is, the model that incorporates all response error parameters (the ε set) and, for the SJ2 task, their accompanying response bias parameters (the κ set). Yet ε parameters are worthy of consideration only if their estimated values differ meaningfully from zero, which, in turn, occurs when the data actually show signs of response errors (e.g., AF responses at long positive auditory delays where only VF responses should occur, or VF responses at long negative auditory delays where only AF responses should occur). García-Pérez and Alcalá-Quintana (2012a) reported evidence to the effect that not all observers commit errors of all types, or commit errors at all. Therefore, we straightforwardly adapted the previously described approach so as to fit each of the models that arise for each task by eliminating all possible subsets of response error parameters. This renders seven additional models for the SJ2 task (three cases in which only one of the ε parameters is fixed at 0, three cases in which two of them are fixed at 0, and one more in which all are fixed at 0), seven additional models for the SJ3 task (identical to those just described, because elimination of an ε parameter also removes its accompanying κ parameter; see the tree diagram in Fig. 4b), and three additional models for the TOJ task (two cases in which only one of the ε parameters is fixed at 0 and another case in which both are fixed at 0; recall that ξ in the TOJ task cannot be removed). Of all the models thus fitted for each task, we selected that for which the Bayesian information criterion (BIC) was lowest. BIC is defined as $G^2 + k\ln(n)$, where k is the number of parameters and n is the number of independent observations. The lower the BIC, the better. Results are reported next for the variant thus selected for each observer, task, and stimulus configuration.

Figure 5 shows data and fitted functions for SJ2, SJ3, and TOJ tasks for 4 participants in the flash–click experiment, also showing summary panels for aggregated data and average fitted curves across the 11 participants. Figure 6 is analogous and for the same participants in the bouncing ball experiment, with summary panels now reflecting data from the 12 participants in this experiment. Data and curves for the remaining participants are given in Section B of the Supplementary Material. Tables 1 and 2 in Appendix 2 list parameter estimates for each participant in each experiment under each task, which also indicate which model variant was selected in each case.

The most salient aspect of Figs. 5 and 6 is the quality of the fit, with curves that accurately follow the path of the data. The chi-square goodness-of-fit test rejected the model in some cases (see stars in Tables 1 and 2), but these rejections are not massive (8 of 69 cases),[4] and, more important, they do not reflect systematic deviations between the path of the data and the path of the fitted curves. Consider the case of observer 3 in the SJ3 task with the flash–click configuration (center panel in the top row of Fig. 5), for whom the model was rejected. Fitted curves follow the path of the data accurately up to Δt = 150 ms (and, actually, across the entire range of auditory delays for AF data), but they cannot accommodate a few stray data points that occur only for VF (blue symbols) and S (red symbols) responses at $\Delta t \geq 200$ ms. It is hard to imagine how a model could produce strangely winding curves that also accommodate these stray data.

These results show that a single model can account for performance on SJ2, SJ3, and TOJ tasks. Fitting the model separately to data from each task allows estimated parameters to differ across tasks, but this does not imply that the resultant differences will be significant or meaningful. A crucial aspect to look into is, thus, whether estimated parameters differ across tasks. Consider first parameter τ. Tables 1 and 2 reveal that estimates are generally negative, implying an auditory advantage (i.e., auditory signals are processed faster than visual signals), in agreement with other sources of evidence on this issue (for a review, see Vroomen & Keetels, 2010). A repeated measures ANOVA with stimulus type (flash–click or bouncing ball) and task (SJ2, SJ3, or TOJ) as factors revealed no significant effects whatsoever: Estimates of τ did not differ significantly across tasks, $F(2, 20) = 1.03$, $p = .375$, or stimulus type, $F(1, 10) = 1.39$, $p = .266$, and there was no interaction either, $F(2, 20) = 0.39$, $p = .679$.

In turn, $1/\lambda_a$ and $1/\lambda_v$ reflect the standard deviations of arrival latencies, which might vary across stimuli but which should not be affected by the task (i.e., by the type of response requested from the observer). It is noteworthy that

[4] The chi-square statistic X^2 was used because it has been shown to be accurate, whereas the empirical Type I error rate of the log-likelihood ratio statistic G^2 nearly doubles its nominal Type I error rate (García-Pérez, 1994; García-Pérez & Alcalá-Quintana, 2012a; García-Pérez & Núñez-Antón, 2001, 2004). On the other hand, if the true rejection rate is .05, the probability of observing eight or more rejections in 69 cases is .022 under the binomial distribution, so that model rejections are only marginally significant (at $\alpha/2$ = .025).

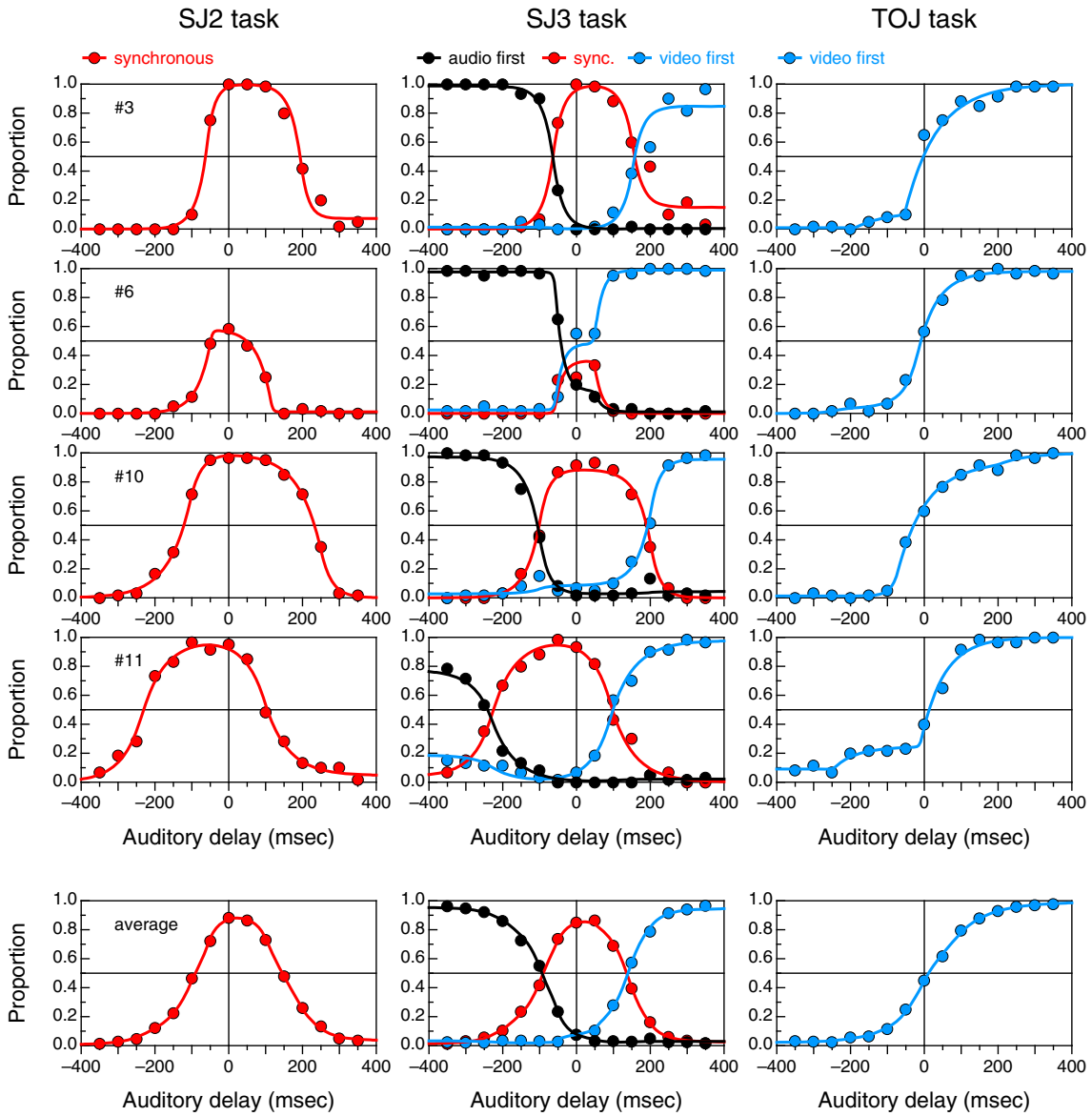

Fig. 5 Data and fitted curves under each task in the flash–click experiment of van Eijk et al. (2008). The upper part shows illustrative results for 4 participants. The bottom part shows summary results as averages of data and averages of the fitted functions across the 11 participants

$1/\lambda_a$ and $1/\lambda_v$ are generally smaller in the bouncing ball than in the flash–click configuration (see Tables 1 and 2). It is also apparent in Tables 1 and 2 that $1/\lambda_a$ and $1/\lambda_v$ are sometimes estimated at 5, which is the lower bound of our parameter space. As is discussed in Appendix 3, this characteristic reflects that the data are not informative of the corresponding parameter and, thus, these boundary values must be regarded as missing data and excluded from analyses. This inevitable decision jeopardizes formal analyses of differences in these parameters across tasks: With repeated measures, all cases with one or more missing values must be discarded. A look at Tables 1 and 2 reveals that this leaves only two cases for analysis. Breaking down the full factorial analysis into separate one-factor analyses does not help much either, because sample size is nevertheless reduced to only 5 observers in some cases. Nevertheless, a comparison of means computed from all available data in each condition is useful. Consider first the case of $1/\lambda_a$. From Table 1, the average ± *SE* estimates across SJ2, SJ3, and TOJ tasks are 42.47 ± 5.54, 36.25 ± 4.96, and 35.21 ± 15.15; from Table 2, they are 20.47 ± 3.10, 25.87 ± 3.07, and 37.82 ± 7.00. The average $1/\lambda_a$ seems to be meaningfully smaller in the bouncing ball than in the flash–click configuration, but, within each configuration, the means do not seem to differ meaningfully across tasks. As for $1/\lambda_v$, Table 1 shows that the average ± *SE* estimates across SJ2, SJ3, and TOJ tasks are 41.58 ± 6.74, 39.33 ± 3.82, and 74.98 ± 12.99; from Table 2, they are 36.32 ± 6.04, 28.37 ± 4.28, and 35.25 ± 12.11. Again, the mean $1/\lambda_v$ seems meaningfully smaller in the bouncing ball than in the flash–click configuration, but, with the only

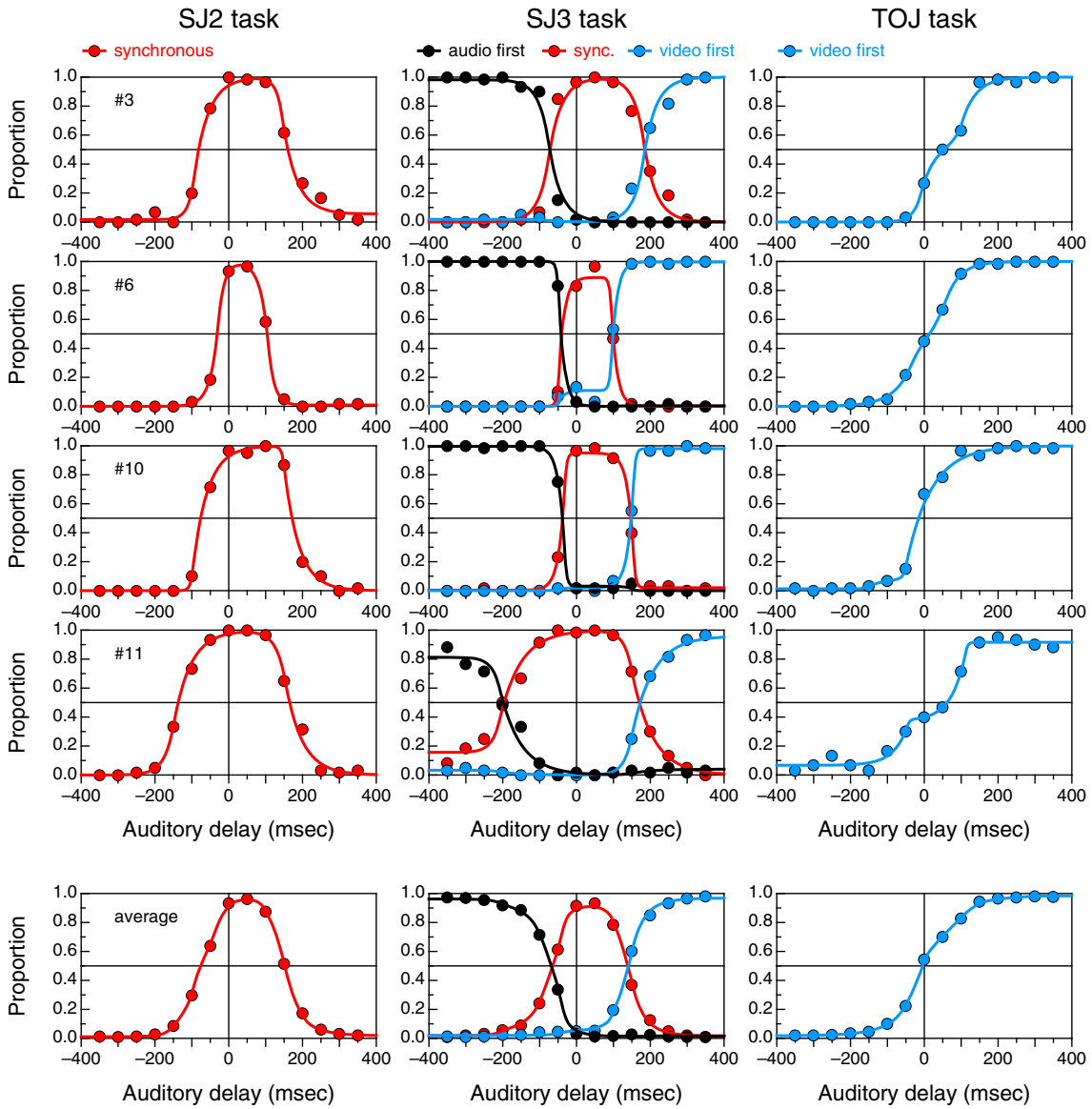

Fig. 6 Data and fitted curves under each task in the bouncing ball experiment of van Eijk et al. (2008). The upper part shows illustrative results for 4 participants (the same as those for whom results in the flash–click experiment were shown in Fig. 5). The bottom part shows summary results as averages of data and averages of the fitted functions across the 12 participants

exception of a remarkably larger mean in the TOJ task with the flash–click configuration, the means do not seem to differ meaningfully across tasks either. We will show below that TOJ tasks render less dependable estimates of these parameters than do SJ2 and SJ3 tasks, which may explain the pattern just described, but another indicator of this problem is the prevalence of boundary estimates of $1/\lambda_a$ and $1/\lambda_v$ from TOJ data (see Tables 1 and 2).

Estimates of δ were also subjected to a repeated measures ANOVA with stimulus type (flash–click or bouncing ball) and task (SJ2, SJ3, and TOJ) as factors, which revealed significant main effects of task only, $F(2, 20) = 12.28$, $p < .001$, without significant main effects of stimulus type, $F(1, 10) = 2.07$, $p = .180$, or interaction, $F(2, 20) = 0.01$, $p = .987$. Subsequent paired comparisons revealed nonsignificant differences between estimates of δ from SJ2 and SJ3 tasks ($p = .651$), whereas estimates from the TOJ task differed significantly from those in the SJ2 task ($p = .010$) and the SJ3 task ($p = .021$). In sum, then, the resolution parameter δ differs significantly across SJ and TOJ tasks.

Finally, parameter ξ is a nuisance brought about by a TOJ task that forces observers to guess for lack of the S response option. Like error and bias parameters (the ε and κ sets), ξ is unrelated to temporal-order perception, and there is no reason to expect any relation of these parameters to stimulus type. In any case, Tables 1 and 2 reveal large between-subjects differences in estimated ξ, with cases that range from bias toward responding VF (i.e., ξ close to 1), bias in

the opposite direction (ξ close to 0), or a more balanced guessing behavior (ξ around .5).

Joint fit

Because the preceding analysis revealed no meaningful or significant differences across tasks in parameters λ_a, λ_v, and τ, we fitted the model jointly to data from the three tasks under the constraint that these parameters have the same values in all tasks, whereas the resolution parameter δ and the error and bias parameters may still differ across tasks. Even if differences in sensory parameters were deemed significant or meaningful across tasks, this joint analysis serves two additional goals: (1) assessing whether the data are compatible with the notion of common sensory parameters across tasks, and (2) using data from all tasks to solve deficiencies that rendered occasional boundary estimates for λ_a and λ_v in the separate analysis of data from each task.

Fitting the model under the joint constraints requires maximizing the joint likelihood function

$$L_{SJ2}(\boldsymbol{R}_{SJ2};\boldsymbol{\theta}_{all}) \times L_{SJ3}(\boldsymbol{R}_{SJ3};\boldsymbol{\theta}_{all}) \times L_{TOJ}(\boldsymbol{R}_{TOJ};\boldsymbol{\theta}_{all}), \qquad (13)$$

where L_{SJ2}, L_{SJ3}, and L_{TOJ} are given by Eqs. 10, 11, and 12, respectively, and where $\boldsymbol{\theta}_{all} = (\lambda_a, \lambda_v, \tau, \delta_{SJ2}, \varepsilon_{SJ2\text{-}AF}, \varepsilon_{SJ2\text{-}S}, \varepsilon_{SJ2\text{-}VF}, \delta_{SJ3}, \varepsilon_{SJ3\text{-}AF}, \varepsilon_{SJ3\text{-}S}, \varepsilon_{SJ3\text{-}VF}, \kappa_{AF\text{-}S}, \kappa_{S\text{-}AF}, \kappa_{VF\text{-}AF}, \delta_{TOJ}, \varepsilon_{TOJ\text{-}AF}, \xi, \varepsilon_{TOJ\text{-}VF})$ is the vector of parameters, where subscripts for δ have been introduced because this parameter can vary across tasks. The likelihood function was maximized as described above for the separate fit. Reduced versions of this full model were also considered as discussed earlier for the separate approach. In this case, an overall number of $8 \times 8 \times 4 = 256$ models resulted from the factorial combination of the eight versions for the SJ2 task (including the full model), eight versions for the SJ3 task, and four versions for the TOJ task. The best-fitting model was also selected as described above.

The results are shown in Figs. 7 and 8 for the same observers as in Figs. 5 and 6 (see Section B of the Supplementary Material for the remaining observers), and Tables 3 and 4 in Appendix 2 list parameter estimates for each observer with each stimulus configuration. The differences with respect to Figs. 5 and 6 are barely noticeable: Both approaches produce curves that pass through the data points almost identically. Interestingly, the joint fit did not result in boundary estimates for parameters λ_a or λ_v (see Tables 3 and 4), a natural outcome because a potential lack of information in the data from any given task is supplemented with information provided by data from other tasks: Regarding the case discussed in depth in Appendix 3, compare the results for observer 6 with the flash–click configuration (second panel down the left column of Figs. 5 and 7). The X^2 statistic rejected the joint model for 5 observers in the flash–click configuration and for 7 in the bouncing ball configuration (see stars in Tables 3 and 4). The reason for this relatively large number of rejections is that the BIC criterion used to select a reduced model penalized models with more parameters that rendered a nonsignificant X^2 statistic.[5] But Figs. 7 and 8 reveal that rejections do not reflect systematic deviations between the path of the data and the path of the fitted curves, even when the model is nominally rejected.

We compared the fit of the separate and joint approaches as follows. First, and because the separate approach entails fitting a separate model to data from each task, the G^2 values from the three separate fits for each observer and stimulus configuration were summed. With this compound G^2 for the separate fit and the G^2 obtained under the joint fit, the BIC for each approach was computed. Figure 9 plots BIC for the compound of the separate fits against BIC for the joint fit. Generally, the BIC of the joint fit is lower, which quantitatively favors this approach, besides the fact that the joint fit with common sensory parameters across tasks renders a more parsimonious and reasonable account of performance when stimuli and conditions are identical across tasks. The almost identical shapes of psychometric functions arising from the separate and joint approaches (compare Figs. 5 and 6 with Figs. 7 and 8) and the lack of significant or meaningful differences in sensory parameters separately estimated for each task (as reported in the preceding section) also support that sensory parameters can be regarded as common across tasks.

To better appreciate differences in model parameters across the two approaches, Fig. 10 plots parameter estimates under the separate approach against parameter estimates under the joint approach for each task. Consider first the left column, for parameter λ_a. Recall that this parameter has the same value for all tasks under the joint approach, which implies that the abscissae of points plotted in all three panels in the left column are the same. In the SJ2 and SJ3 tasks (top and center panels in that column), this parameter has a similar value whether estimated under the separate (ordinate) or joint (abscissa) approaches. In contrast, the value of this parameter as estimated separately from TOJ data is barely related to its estimated value under the joint approach (bottom panel in the first column of Fig. 10). The same holds for parameters λ_v (center-left column in Fig. 10) and τ (center-right column of Fig. 10), which also have the same value across tasks under the joint approach and which attain similar values when SJ2 or SJ3 data are fitted separately but very disparate values when TOJ data are fitted separately. The right column of Fig. 10, for parameter δ, represents a slightly different situation, because this parameter differs

[5] In contrast, for the simpler (separate) models considered earlier, the model yielding the lowest BIC was also always the model that best fitted the data according to the X^2 statistic.

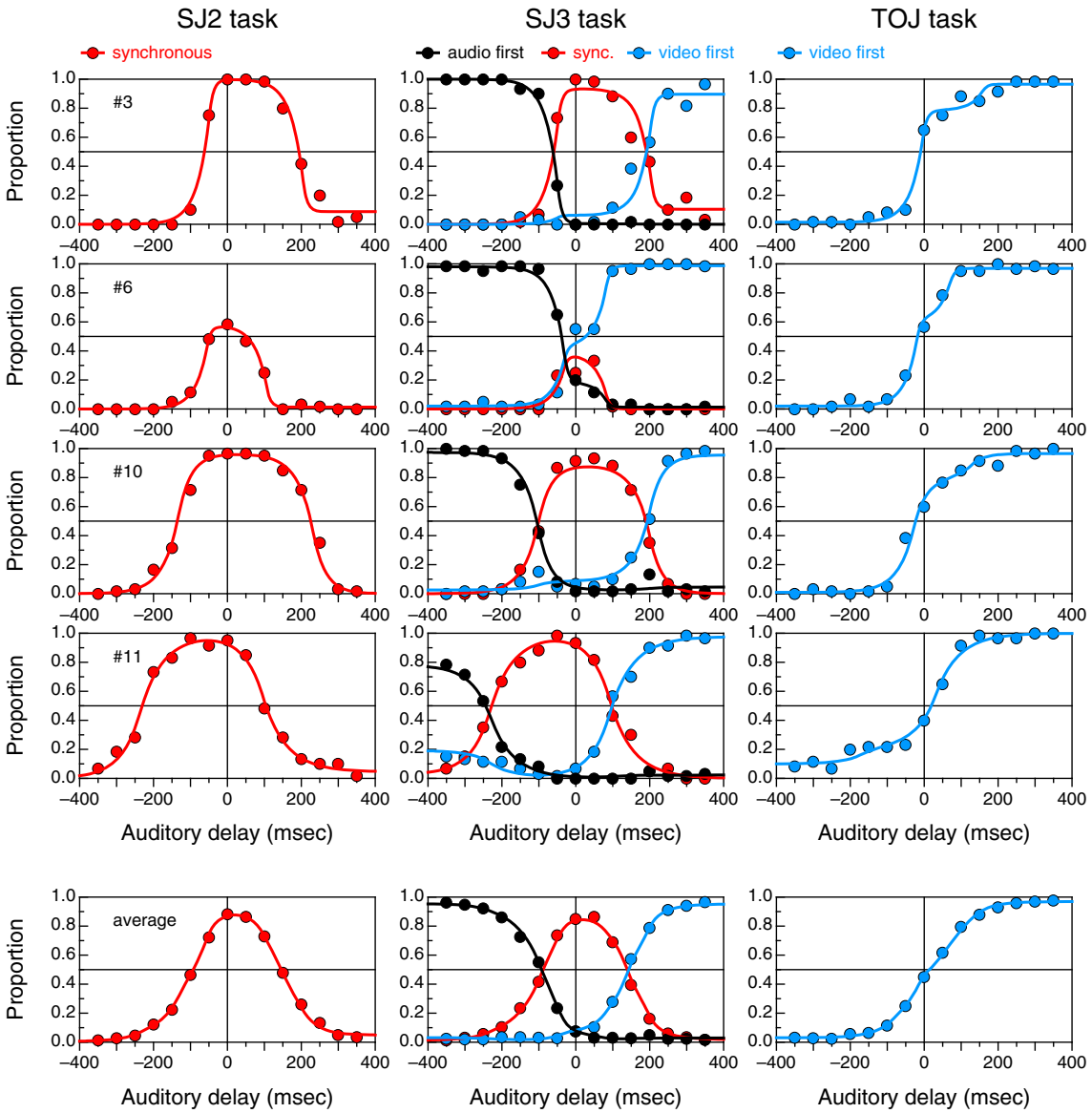

Fig. 7 Data and curves fitted jointly to the three tasks in the flash–click experiment of van Eijk et al. (2008). If the fitted curves plotted here were plotted in Fig. 5, the two sets of curves would superimpose in almost all cases. Graphical conventions as in Fig. 5

across tasks also under the joint approach. Nevertheless, the joint and separate estimates of δ are nearly identical for SJ2 and SJ3 tasks (top and center panels in the right column of Fig. 10), and, again, they differ more for TOJ tasks (bottom panel in that column).

The reason for the large discrepancy between separate and joint estimates for TOJ data lies in a subtle aspect of the task. TOJ tasks provide only one source of data (VF responses at each auditory delay) that embodies an inextricable mixture of true judgments, misreports, and guesses. The mixture renders likelihood functions with multiple local maxima and, hence, offers a multiplicity of exit points to a parameter estimation algorithm whose termination criterion is simply based on the change across iterations. The result is inaccurate parameter estimates with large standard errors (see Fig. 14 below). Under the joint approach, in contrast, parameter estimates that describe TOJ data are sought using the independent evidence provided by data from SJ2 and SJ3 tasks, rendering a solution similar (in the maximum-likelihood sense) to that obtained under the separate fit of TOJ data but in which parameters λ_a, λ_v, and τ are not permitted to vary wildly (compare model curves in the right columns of Figs. 5 and 7, which are very similar despite the different parameter values).

Analyses of parameter estimates under the joint approach rendered conclusions that corroborate those of the separate approach. Estimates of τ were larger in absolute value in the bouncing ball than in the flash–click configuration (see Tables 3 and 4), but a paired-samples *t*-test revealed non-significant differences, $t(10) = 1.36$, $p = .202$. For estimates

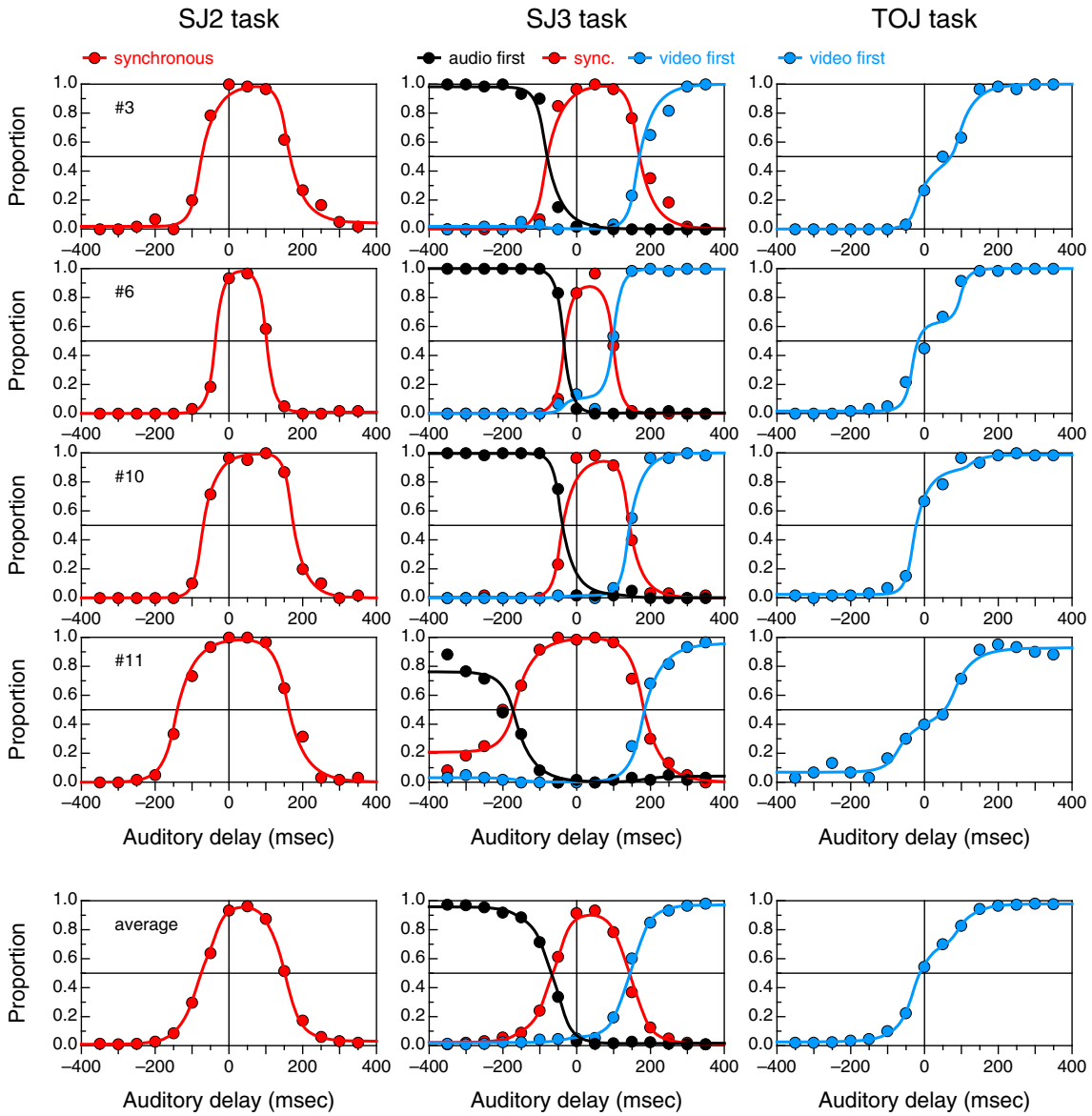

Fig. 8 Data and curves fitted jointly to the three tasks in the bouncing ball experiment of van Eijk et al. (2008). If the fitted curves plotted here were plotted in Fig. 6, the two sets of curves would superimpose in almost all cases. Graphical conventions as in Fig. 6

of λ_a and λ_v, a repeated measures ANOVA with modality (auditory or visual) and stimulus type (flash–click or bouncing ball) as factors yielded only significant main effects of stimulus type, $F(1, 10) = 9.46$, $p = .012$, with no differences between $1/\lambda_a$ and $1/\lambda_v$, $F(1, 10) = 0.87$, $p = .374$, and no interaction, $F(1, 10) = 0.64$, $p = .442$. As for estimates of δ, a repeated measures ANOVA with stimulus type (flash–click or bouncing ball) and task (SJ2, SJ3, and TOJ) as factors yielded only significant main effects of task, $F(2, 20) = 33.51$, $p < .001$, with no main effects of stimulus type, $F(1, 10)=2.48$, $p=.146$, and no interaction, $F(2, 20) = 0.77$, $p = .475$. Subsequent paired comparisons revealed nonsignificant differences between estimates of δ from SJ2 and SJ3 tasks ($p = .998$), whereas estimates from the TOJ task differed significantly from those in the SJ2 task ($p < .001$) and the SJ3 task ($p < .001$).

Conventional indices of performance

Van Eijk et al. (2008) estimated synchrony ranges and PSSs by fitting piecewise cumulative and survival Gaussians to S responses, cumulative Gaussians to VF responses, and survival Gaussians to AF responses. Our approach (which fits model-based functions instead) differs from theirs, but it is unlikely that the two approaches yield meaningfully different performance measures when the alternative curves used to obtain

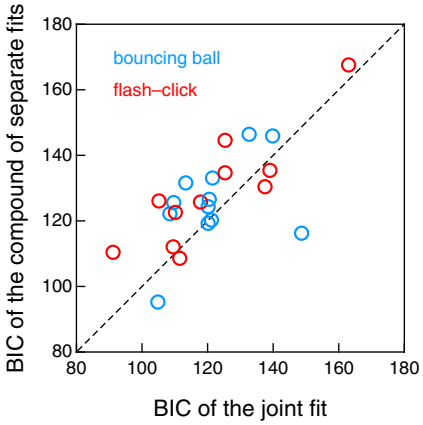

Fig 9 Scatterplot of BIC for the compound of the separate fits against BIC for the joint fit, in the flash–click and the bouncing ball experiments. Each data point pertains to a different observer

them describe the skeleton of the data similarly. We nevertheless estimated PSSs from our fit using Eqs. 17, 18a, 18b, 19a, 19b, 20a, 20b, 21a, 21b, 22a, 22b and 22c in Appendix 1 after removing response error and bias parameters as in van Eijk et al. (2008, 2010).

As was expected, curves estimated from our separate approach (Figs. 5 and 6) rendered performance indices that corroborated the conclusions of van Eijk et al. (2008). For the flash–click stimulus, average ± *SE* PSSs for SJ2, SJ3, and TOJ tasks were, respectively, 24.12 ± 9.88, 15.91 ± 8.89, and 14.71 ± 12.27, the correlation between PSS estimates from SJ2 and SJ3 tasks was .942 ($p < .001$), and correlations between PSS estimates from SJ2 and TOJ tasks ($r = -.166$) or SJ3 and TOJ tasks ($r = .098$) were not significant. For the bouncing ball stimulus, average ± *SE* PSSs for SJ2, SJ3, and TOJ tasks were, respectively, 36.33 ± 6.20, 32.01 ± 5.83, and 7.39 ± 9.55, the correlation between PSS estimates from SJ2 and SJ3 tasks was .771 ($p = .003$), and no significant correlation was found between estimates from SJ2 and TOJ tasks ($r = -.249$) or SJ3 and TOJ tasks ($r = -.145$). We should point out that we did not need to exclude observer 6 in the flash–click experiment from our analyses, because our

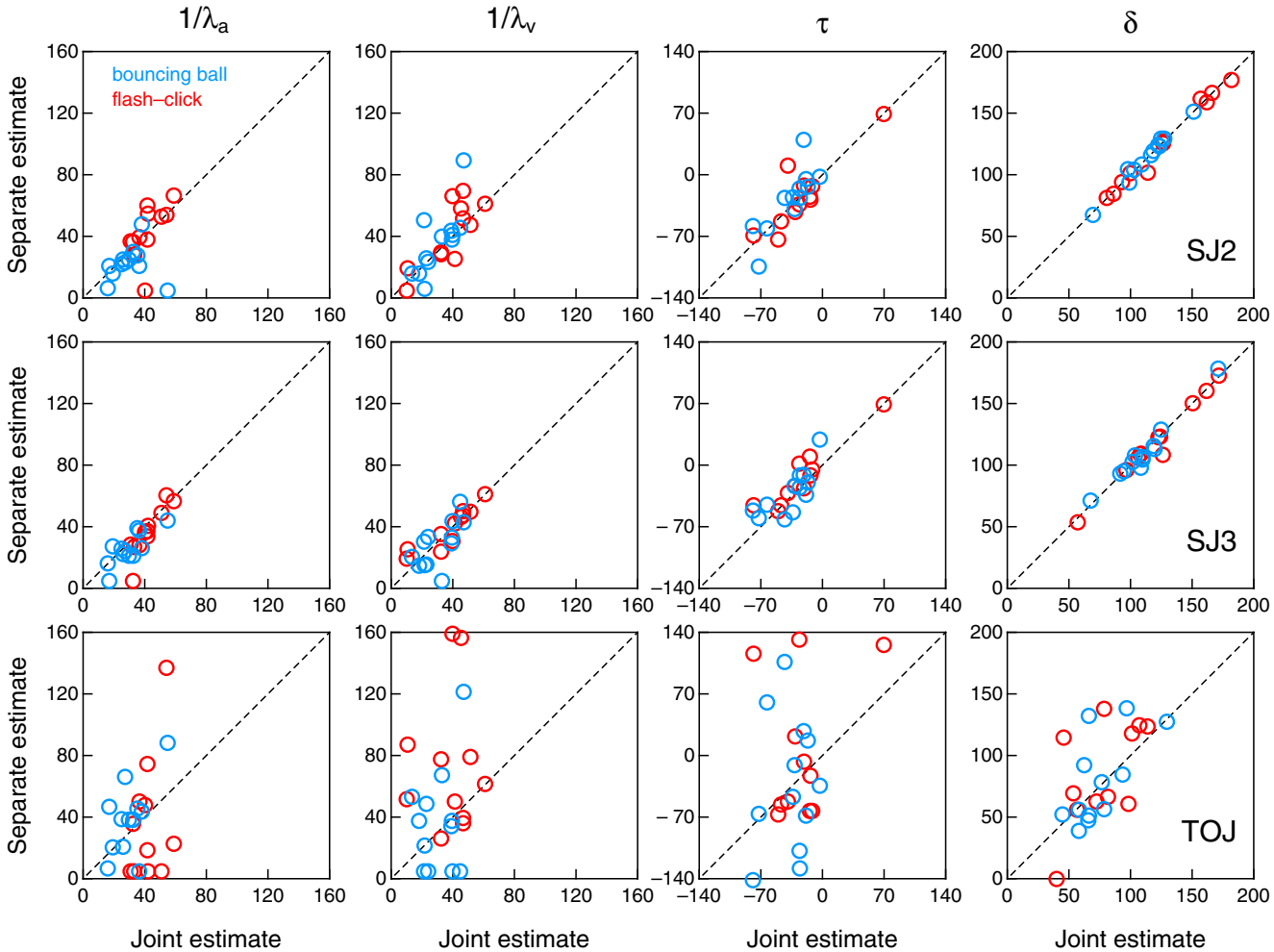

Fig. 10 Scatterplots of parameters estimated under the joint approach (horizontal axes) and parameters estimated under the separate approach (vertical axes). Each row pertains to one of the three tasks (from top to bottom, SJ2, SJ3, and TOJ), and each column pertains to one of the relevant parameters (left to right, $1/\lambda_a$, $1/\lambda_v$, τ, and δ). Each data point comes from a different observer

fitted curves in the SJ3 task cross and, thus, allow estimating a PSS.

Repeating the analyses with parameters from our joint approach (Figs. 7 and 8) rendered performance measures that were almost identical across SJ2 and SJ3 tasks. For the flash–click stimulus, average ± *SE* estimates of the PSS were 21.33 ± 9.62 in the SJ2 task and 21.36 ± 9.61 in the SJ3 task, whereas, for the bouncing ball stimulus, they were 36.45 ± 5.53 and 36.43 ± 5.52. For both stimuli, the correlation between PSS estimates from SJ2 and SJ3 tasks was in excess of .9999. As for the TOJ task, average ± *SE* estimates of the PSS were 13.36 ± 13.35 with the flash–click stimulus and −0.89 ± 12.02 with the bouncing ball stimulus. The correlations between TOJ and SJ2 estimates of the PSS were virtually null (−.051 and − .028 with the flash–click and the bouncing ball stimuli, respectively), and so were the correlations between TOJ and SJ3 estimates (−.051 and − .027 with the flash–click and the bouncing ball stimuli, respectively).

Van Eijk et al. (2008) also analyzed a sensitivity measure defined as the slope of the fitted TOJ curve at the PSS. For the cumulative Gaussians fitted to TOJ data by van Eijk et al. (2008), these slopes relate to the standard deviation of the Gaussian and are useful for describing the spread of the entire curve. In contrast, our model renders curves whose slope may not change smoothly (see the right column of Figs. 5 and 6), and, then, the slope at any arbitrary point gives a very poor description of the spread of the entire curve. A way around this difficulty is to measure spread directly with a method akin to that used to define comparable measures of the spread of psychometric functions across variations in their mathematical form (Alcalá-Quintana & García-Pérez, 2004; García-Pérez, 1998; García-Pérez & Alcalá-Quintana, 2005; Gilchrist, Jerwood, & Ismaiel, 2005) or to that used to define comparable measures of variability across variations in the form of distributions (Townsend & Colonius, 2005). Because the standard deviation of a cumulative Gaussian reflects half the distance between the 15.87 % and the 84.13 % points on the curve, a comparable measure for our model curves is also half the distance between these points on $\Psi_{\text{TOJ-VF}}$. These points, $\Delta t_{\text{TOJ-.1587}}$ and $\Delta t_{\text{TOJ-.8413}}$, are readily obtained through Eq. 22a, 22b and 22c in Appendix 1, and the spread of $\Psi_{\text{TOJ-VF}}$ is $\sigma_{\text{TOJ}} = (\Delta t_{\text{TOJ-.8413}} - \Delta t_{\text{TOJ-.1587}})/2$.

With parameter values in Tables 1 and 2 (from our separate approach), the average ± *SE* spread was 101.47 ± 13.07 with the flash–click stimulus and 71.14 ± 11.05 with the bouncing ball stimulus, implying a narrower spread (i.e., a higher sensitivity) with the latter stimulus. Results using parameter values from our joint approach (Tables 3 and 4) were similar owing to the fact that the shapes described by the curves from which these measures are obtained are very similar under both approaches. But van Eijk et al. (2008; see their Table 2) reported slopes instead of spread. By defining the *putative* slope at the 50 % point on our curves as $1/\left(\sqrt{2\pi}\sigma_{\text{TOJ}}\right)$ (i.e., through the same transformation used to obtain the slope at the 50 % point on a cumulative Gaussian), we obtained average ± *SE* slopes of $0.47 \times 10^{-2} \pm 0.06 \times 10^{-2}$ in the flash–click condition and $0.68 \times 10^{-2} \pm 0.08 \times 10^{-2}$ in the bouncing ball condition, in good agreement with the results that van Eijk et al. (2008) reported.

Recall, however, that sensitivity measures from TOJ tasks can be overly misleading because they are contaminated by the response bias of the observer. This contaminating influence is best appreciated when δ is large, in which case the spread of $\Psi_{\text{TOJ-VF}}$ is very broad unless the irrelevant response bias parameter ξ is almost 0 or almost 1. The bottom panel of Fig. 3 illustrates this characteristic: The 15.87 % point on $\Psi_{\text{TOJ-VF}}$ occurs at a very long negative auditory delay when ξ = .2 (continuous gray curve) but at a very long positive auditory delay when ξ = 0 (blue curve), while in both cases, the 84.13 % point is almost at the same location. The consequences on sensitivity measures are dramatic. It is thus unsurprising that sensitivity measures from TOJ tasks do not agree well with purportedly analogous sensitivity measures from SJ2 or SJ3 tasks.

In sum, in terms of conventional sensitivity measures, synchrony ranges, and estimates of the PSS, our fitting approach and that of van Eijk et al. (2008) yield essentially the same conclusions, although a simple analysis of performance measures cannot elucidate how SJ and TOJ tasks differ in terms of the processes governing performance in each task.

Discussion

SJ and TOJ tasks are widely used in studies on prior-entry effects. As Spence and Parise (2010) put it, "understanding the effect has been hindered by the many methodological confounds present in early research. As a consequence, it is unclear whether the behavioral effects reported in the majority of published studies in this area should be attributed to attention, decisional response biases, and/or, in the case of exogenous spatial cuing studies of the prior-entry effect, to sensory facilitation effects instead" (p. 364). They also acknowledged that prior-entry effects tend to be smaller when performance is assessed with SJ tasks than when assessed with TOJ tasks, and they also discussed the problems and potential biases inherent in either task. Yarrow et al. (2011) have also discussed and demonstrated these confounds empirically.

Under the usual practice of fitting arbitrary functions to the data, the estimated parameters do not speak of the processes under study—namely, whether arrival latencies of attended stimuli are shorter than those of unattended stimuli. In contrast, the model considered here includes

parameters that represent the distributions of these arrival latencies along with other parameters that reflect decisional aspects of the task, response bias, and response errors. By fitting these model-based functions to data, the estimated parameters directly speak of the underlying processes. We have shown that the model gives a satisfactory account of the data of van Eijk et al. (2008), and an analysis of parameter estimates across tasks and stimulus configurations revealed aspects that conventional Gaussians fitted to the data cannot disclose. Specifically, (1) the distributions of arrival latencies do not differ across SJ2, SJ3, and TOJ tasks, but the resolution parameter δ does differ across tasks; (2) arrival latencies have a smaller standard deviation (i.e., $1/\lambda_a$ and $1/\lambda_v$ are lower) with the bouncing ball stimulus than with the flash–click stimulus, surely as a result of the anticipatory information provided by the former; (3) the auditory advantage reflected in parameter τ is larger (although not significantly) with the bouncing ball stimulus than with the flash–click stimulus, perhaps for the same reason; and (4) the observers' resolving power (i.e., the ability to discriminate small differences in arrival latency, as indicated by parameter δ) is also higher with the bouncing ball stimulus, perhaps because the availability of anticipatory information allows observers to better allocate their attentional resources and perform at a level closer to their absolute resolution limit.

The data analyzed in this article are compatible with the notion that sensory parameters are common to all tasks when stimulus conditions are identical, whereas decisional parameters differ across tasks. But some assumptions, implications, extensions, and consequences of the model are worth discussing.

General applicability of the model

Independent-channels models embody a representation of the processes governing judgments of temporal order or simultaneity, and our addition of response-error parameters account for empirical evidence of misreports in SJ or TOJ tasks. All the parameters in our model have empirical referents; hence, their inclusion is justifiable. We have used the data of van Eijk et al. (2008) in our test and illustration of the model because theirs seems to be the only experiment carried out using the same observers and conditions under the three tasks. But there is nothing peculiar about this data set in the sense that data reported in other studies and collected only under two of the tasks display similar characteristics (e.g., Donohue et al., 2010; Fujisaki & Nishida, 2009; Schneider & Bavelier, 2003; van Eijk et al., 2010; Vatakis et al., 2008; Yates & Nicholls, 2011). Analyses of other data sets that allow addressing additional issues will be reported in due course.

A characteristic of the experiment whose data we reanalyzed is that attentional conditions were not manipulated in any way beyond what the flash–click or bouncing ball conditions bring around (anticipatory information about the time of stimulus presentation in the latter case, as compared with an unpredictable time of stimulus presentation in the former). The question thus may arise as to whether the model would also hold under the conditions of studies on prior entry or temporal recalibration, in which several manipulations are involved. Although the ultimate answer to this question must wait until data from those studies are analyzed under our model, the empirical patterns of results reported in such experiments (e.g., Schneider & Bavelier, 2003; Stelmach & Herdman, 1991; Yates & Nicholls, 2011) are certainly compatible with our model.

On another front, misreports can have causes other than mere errors in pressing the response keys. Yamamoto and Kitazawa (2001; see also Shore et al., 2002) had their participants carry out a TOJ task to indicate which of their two hands had received a mechanical stimulation earlier. Their results showed that the psychometric function for "right-hand earlier" responses had the typical sigmoidal and monotonic shape when the experiment was carried out with the hands in their natural position; but with crossed hands, the function was N-shaped. They argued that short arrival-time differences cause misreports of temporal order when arms are crossed. In contrast to unconditional response errors in our model, this factor causes misreports that do not have a fixed probability across arrival-time differences; rather, their probability is largest at null arrival-time differences and decreases as arrival-time difference increases in absolute value. An extension of our model along these lines describes Yamamoto and Kitazawa's results, but details are omitted here.

Distribution of arrival latencies

Our model is based on exponential distributions of arrival latencies. One might consider using alternative and, presumably, more realistic distributions given by the gamma densities

$$g_i(t) = \frac{\lambda_i^{\alpha_i}}{\Gamma(\alpha_i)}(t - (\Delta t_i + \tau_i))^{\alpha_i - 1}\exp[-\lambda_i(t - (\Delta t_i + \tau_i))],$$
$$t \geq \Delta t_i + \tau_i, \qquad (14)$$

of which our exponential distributions are particular cases with $\alpha_i = 1$.

This simple replacement makes the model intractable because the arrival-time difference D then has a bilateral gamma density that has no closed-form expression for arbitrary values of the shape parameter α (Holm & Alouini, 2004; Küchler & Tappe, 2008), a problem that carries over to the cumulative distributions needed to define the psychometric functions for SJ2, SJ3, and TOJ tasks. Parameter

estimation thus becomes a formidable computational task. Yet Fig. 11 shows through a tractable case that bilateral gamma densities have shapes similar to those of bilateral exponential densities (compare with Fig. 1; see also Fig. 1 in Küchler & Tappe, 2008). Although the data of van Eijk et al. (2008) can be accurately accounted for with exponential distributions, there is room for replacing them with other realistic (i.e., causal) distributions if those are mathematically tractable.

The relation between detection of asynchrony and identification of temporal order

Two recurring issues in perception of temporal order are whether the detection of asynchrony is a sufficient condition for the identification of temporal order and whether detection of synchrony and identification of temporal order arise from a single processing stream (e.g., Allan, 1975; Hirsh & Sherrick, 1961; Mitrani, Shekerdjiiski, & Yakimoff, 1986; see also Occelli, Spence, & Zampini, 2011; Shore et al., 2002). The three-zone decision space in independent-channels models (see Fig. 1b) assumes that observers' judgments are ternary outcomes, not the result of two independent binary decisions. The only model of the latter type seems to be that of Jaśkowski (1991a), in which two independent paths are invoked: one to judge successiveness and the other to judge temporal order. The model also includes

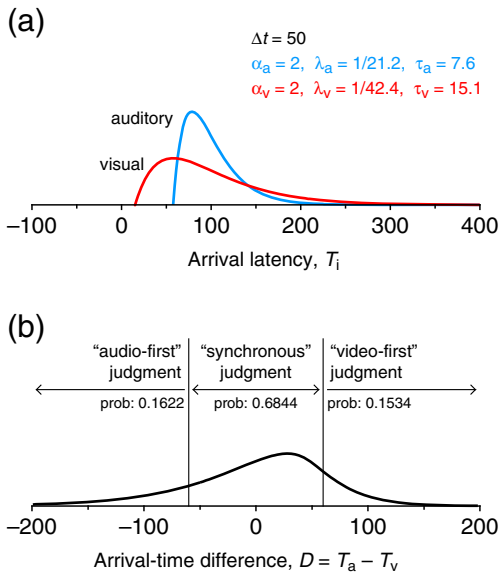

Fig. 11 **a** Gamma distributions for the arrival latency of a visual stimulus (red curve) presented at time 0 and an auditory stimulus (blue curve) presented at time $\Delta t = 50$ ms. Parameters indicated in the inset are such that the mean and standard deviation of arrival latencies of visual and auditory stimuli attain the same values as they had under the exponential distributions in Fig. 1a. **b** Bilateral gamma density for the arrival-time differences, which resembles a smoothed version of the bilateral exponential density in Fig. 1b

guessing mechanisms to solve cases in which the outcomes of the two independent paths are incompatible or inconclusive, and Jaśkowski (1991a, 1991b, 1993) showed that this model can fit data. Because independent-channels models and two-stage models offer diametrically different accounts of SJs and TOJs, a discussion of the empirical evidence supporting either account is worth considering.

Allan (1975) used a dual-response task in which, on each trial, observers first reported whether or not the offsets (instead of the onsets) of a visual and an auditory stimulus were simultaneous and, subsequently and regardless of their first response, whether the offset of the visual or the auditory stimulus preceded the other. By collecting a dual response on each trial, Allan could demonstrate that detection of asynchrony is a necessary and a sufficient condition for identification of temporal order, thus favoring the account offered by independent-channels models. In contrast, Mitrani et al. (1986) claimed to report evidence that detection of asynchrony is not sufficient for the identification of temporal order, which would thus favor two-stage models. Yet their data do not justify their conclusion, for various reasons. First, instead of a dual-response task, Mitrani et al. used an SJ3 task that does not allow elucidating whether or not a judgment of asynchrony is always followed by a report of temporal order. Second, and more important, their conclusion arises from an incorrect interpretation of the data. In particular, they computed the frequency of each type of response ("left first," "right first," or "simultaneous," since their visual stimuli were two LEDs horizontally aligned) at various SOAs, and they defined the "prevalent" response as that which was given significantly more frequently than the two other responses. They also assumed that observers actually (and only) perceived whatever this prevalent response was. Their conclusion that, at short SOAs, "the subjects were not able to indicate the order of flash presentation although they did not perceive them as simultaneous" (p. 161) is supported only by the observed lack of prevalent responses: "Simultaneous" responses did not prevail (which was taken as indicating that simultaneity was not perceived) and "left first" or "right first" responses also did not prevail (which was taken as evidence that temporal order was not perceived either). A more fitting interpretation of their data is that observers perceived simultaneity, left first, or right first on different trials and with no overall prevalence. Independent-channels models can account for this outcome.

Measures of latent sensitivity and latent PSS

Parameter δ reflects the observer's ability to resolve small differences in arrival latency. Given that this parameter affects all performance measures (see Eqs. 17, 18a, 18b,

19a, 19b, 20a, 20b, 21a, 21b, 22a, 22b and 22c in Appendix 1), it is useful to define *latent performance* as the psychometric function for VF judgments if the observer had infinite resolution (i.e., as if $\delta = 0$), also setting all response error and bias parameters to 0. These *latent* psychometric functions describe performance as limited only by the random nature of arrival latencies and are shown in Fig. 12 for each observer, using estimates of λ_a, λ_v, and τ from Tables 3 and 4. These functions can also be used to define alternative and "pure" estimates of the PSS (the 50 % point on the curve) and sensitivity (spread defined as the distance between the 84.13 % and the 15.87 % points on the curve). The *latent PSS* is given by

$$\theta = \begin{cases} \frac{1}{\lambda_a}\ln\left[\frac{\lambda_a+\lambda_v}{2\lambda_v}\right] - \tau & \text{if } \lambda_v \geq \lambda_a \\ \frac{1}{\lambda_v}\ln\left[\frac{2\lambda_a}{\lambda_a+\lambda_v}\right] - \tau & \text{if } \lambda_v < \lambda_a \end{cases} \quad (15)$$

and the *latent sensitivity* (spread) is given by

$$\sigma = \begin{cases} \frac{1}{\lambda_a}\ln\left[\frac{0.8413}{0.1584}\right] & \text{if } \frac{\lambda_v}{\lambda_a+\lambda_v} \geq 0.8413 \\ \frac{1}{\lambda_v}\ln\left[\frac{0.8413}{0.1584}\right] & \text{if } \frac{\lambda_v}{\lambda_a+\lambda_v} \leq 0.1587 \\ \frac{\ln\lambda_a}{\lambda_v} + \frac{\ln\lambda_v}{\lambda_a} - \left(\frac{1}{\lambda_v}+\frac{1}{\lambda_a}\right)\ln[0.1587(\lambda_a+\lambda_v)] & \text{otherwise} \end{cases}. \quad (16)$$

Figure 13 plots latent PSSs and sensitivities across stimulus configurations, showing that the PSS tends to be a little higher and that sensitivity is clearly higher (spread is smaller) with the bouncing ball stimulus. Note that this measure of sensitivity is purely determined by sensory factors (arrival latencies), whereas the resolution measured by parameter δ additionally limits performance in a way that can vary across tasks, attentional conditions, or other factors. Thus, the triplet (θ, σ, $\pm\delta$) gives latent performance measures in which sensory factors are separated from decisional factors, allowing separate assessment of decisional influences (parameter δ) and stimulus/attentional influences (parameters θ and σ, which depend only on λ_a, λ_v, and τ). Assessing performance in this way may also be useful in research on neuropsychological conditions in which temporal-order judgments are impaired. For instance, dyslexics find difficulty discriminating temporal order, and it would be useful to know whether these difficulties have a sensory or a decisional basis, given their effect on performance at visual detection and discrimination tasks (Ben-Yehudah, Sackett, Malchi-Ginzberg, & Ahissar, 2001; Landerl & Willburger, 2010; Peli & García-Pérez, 1997; Ram-Tsur, Faust, & Zivotofsky, 2006).

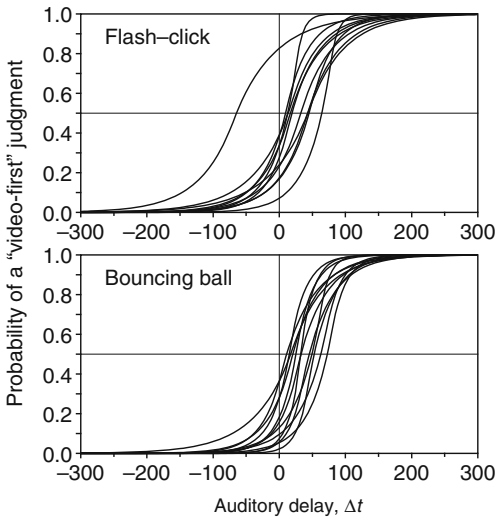

Fig. 12 Latent psychometric functions for VF judgments in the flash–click and bouncing ball experiments. Each curve pertains to a different observer

Which task?

Although conducting experiments using the three tasks as done by van Eijk et al. (2008) is useful and informative, research is more efficient when carried out with a single task. In principle, all tasks are eligible because fitting model-based psychometric functions provides estimates of the same parameters. The question, then, involves

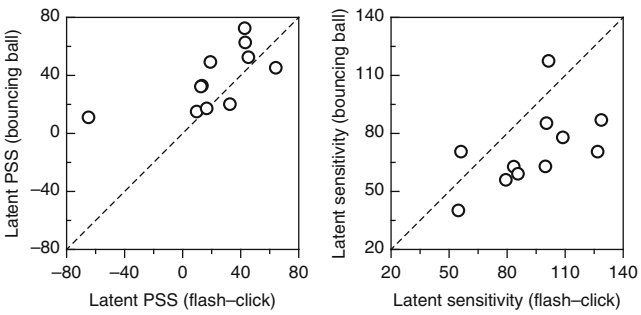

Fig. 13 Scatterplot of measures of latent PSS (50 % point on the latent psychometric functions in Fig. 12) and latent sensitivity (spread of the same latent psychometric functions) in the bouncing ball experiment against the same measures in the flash–click experiment. Each data point pertains to a different observer

consideration of which task yields more dependable estimates. To address this question, we used simulation methods analogous to those in García-Pérez and Alcalá-Quintana (2012a), except that all error parameters were assumed to be zero here, so as to obtain an uncontaminated picture of the extent to which the tasks themselves allow recovering parameters. Thus, 1,000 data sets were produced under randomly drawn parameters with uniform distributions on

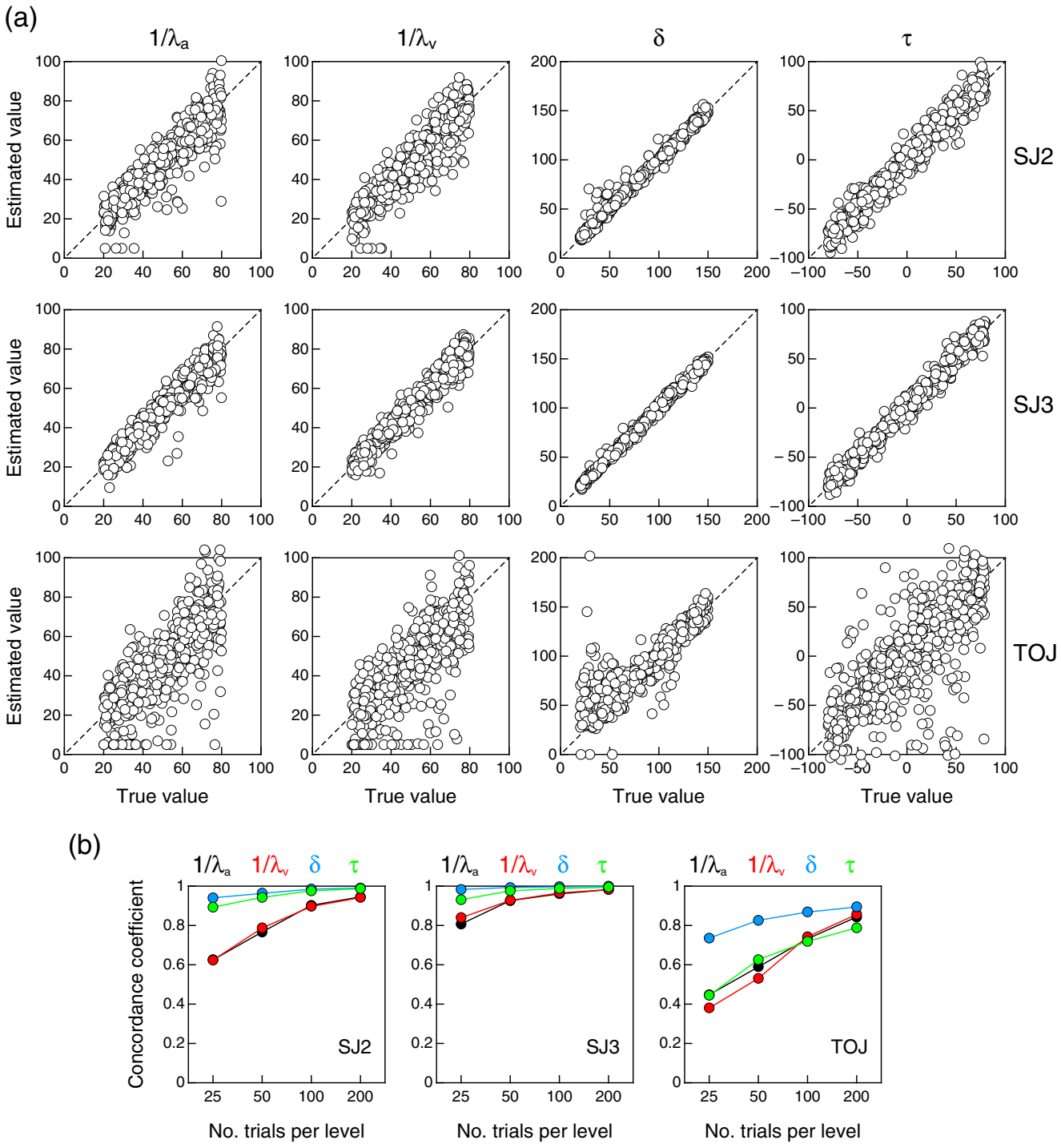

Fig. 14 **a** Scatterplot of estimated values against true values of the most relevant model parameters (column headers) from isolated SJ2, SJ3, and TOJ tasks (rows) with 100 trials at each of 15 auditory delays. **b** Summary measures of agreement between estimated and true values of the parameters as a function of number of trials per auditory delay under each task (panels)

[1/80, 1/20] for λ_a and λ_v (independently), on [−80, 80] for τ, on [20, 150] for δ, and on [.1, .9] for ξ (for the TOJ task). The particular parameters that would produce the data set under each task were inserted into Eqs. 7a, 7b, 7c, 8a, 8b, 9a and 9b (as applicable), and responses at auditory delays ranging from −350 to 350 ms in steps of 50 ms were simulated for various numbers of trials per auditory delay. The full and reduced models for each individual task were subsequently fitted to the applicable data from each simulee, and the model with the lowest BIC was selected.[6] The agreement between true and estimated parameters was measured for each task under each number of trials per auditory delay. Figure 14a shows scatter plots of estimated against true parameter values under each task (rows) for the case of 100 trials per auditory delay, where it is clearly apparent that TOJ tasks render larger discrepancies between estimated and true parameters. Figure 14b shows how estimation accuracy varies with number of trials under each task as indicated by the coefficient of concordance (Lin, 1989), which measures the scatter of data around the identity line. It is evident that SJ3 tasks yield slightly better parameter recovery than do SJ2 tasks at small numbers of trials and that TOJ tasks yield much poorer results even with large numbers of trials.

This picture deteriorates when response errors occur, which means only that more data (i.e., more presentations per auditory delay) are needed to compensate for the effect of errors. Another way to ensure accurate parameter recovery when errors are likely to occur is to repeat the experiment with two or three tasks and then fit the model jointly to all data. For simulation results addressing this issue, see Alcalá-Quintana and García-Pérez (2012). But the analyses of empirical data reported here additionally demonstrate that collecting data with more than the SJ3 task is barely useful, even when response errors occur. Note, in the center row of Fig. 10, that parameter estimates from the separate fit to SJ3 data are very similar to estimates arising from the joint fit to data from all tasks, whereas the relation is weaker across the board for separate estimates from the SJ2 task (top row in Fig. 10). This means that use of a lone SJ3 task yields parameter estimates analogous to those obtained by taking the extra step of repeating the experiment with the two other tasks and estimating parameters under the joint approach. The reason for this is surely that the ternary SJ3 task provides sufficient diversification of responses to allow an accurate estimation of parameters without the help of additional data from other tasks. And note that concerns about contamination arising from the observer's criterion (parameter δ) to give S responses should dissipate because this parameter is estimated and does not contaminate the remaining parameters.

Yet, and despite its shortcomings, TOJ tasks seem to place observers in a situation in which they push their resolution limit, given that data from these tasks rendered the lowest estimates of δ in our analyses. If the aim is to find out where this limit is, TOJ tasks are useful, although they should be complemented with SJ2 or SJ3 tasks to obtain dependable estimates of all model parameters.

Conclusion

We have considered a quantitative model that places performance in SJ2, SJ3, and TOJ tasks under a common framework with interpretable parameters and incorporating response errors. The model fits data reported by van Eijk et al. (2008) adequately. An analysis of parameter estimates revealed that differences across tasks are a consequence of differences in the decisional factor represented by parameter δ. Separation of decisional and sensory aspects also allows obtaining alternative and task-independent indices of latent performance in temporal-order judgments.

A formal analysis of the model reveals the shortcomings of TOJ tasks, which arise from the contaminating response bias of observers who are forced to guess when they would have, instead, given the (not allowed) S response. This influence contaminates performance measures from TOJ tasks, which thus lose their relation to analogous measures from SJ2 or SJ3 tasks. We have also shown that lack of bias in TOJ tasks (i.e., $\xi = .5$) does not solve the problem. This is reminiscent of a similar contamination that has the same deleterious effect in 2AFC tasks aimed at estimating psychometric functions and thresholds in detection or discrimination experiments and that can only be surmounted by replacing the binary response format of the 2AFC task with a ternary format in which observers are additionally allowed to give an "I don't know" response (Alcalá-Quintana & García-Pérez, 2011; García-Pérez, 2010; García-Pérez & Alcalá-Quintana, 2010a, 2010b, 2011a, 2011b, 2012b; García-Pérez, Alcalá-Quintana, Woods, & Peli, 2011). These results concur with those of Ulrich (1987) and van Eijk et al. (2008) in advising against the use of TOJ tasks in research on perception of temporal order (see also Spence & Parise, 2010).

User-friendly software packages (in MATLAB and R) have been developed for fitting SJ and TOJ data either separately

[6] It is worth noting that the BIC selected the wrong model with rates that varied between 3 % and 19 % across simulations involving different sample sizes and tasks. In contrast, a concurrent .05-size chi-square test rejected the correct model with rates that varied between 3 % and 7 %.

or jointly, as was done in this article, and under alternative sets of assumptions about how response errors occur (Alcalá-Quintana & García-Pérez, 2012).

Acknowledgments This research was supported by grant PSI2009-08800 from Ministerio de Ciencia e Innovación (Spain). We thank Hans Colonius for calling our attention to research on timing judgments and Rob van Eijk for kindly making their data available for our analyses. We also thank Keith Schneider and two anonymous reviewers for their comments on an earlier draft of the manuscript. Correspondence concerning this article should be sent to Miguel A. García-Pérez, Departamento de Metodología, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain (e-mail: miguel@psi.ucm.es).

Appendix 1

This appendix shows how conventional indices of performance typically obtained from SJ2, SJ3, and TOJ data are related to model parameters.

First note that the psychometric function $\Psi_{SJ3\text{-}S}$ for "synchronous" judgments in SJ3 tasks (which is identical to $\Psi_{SJ2\text{-}S}$, as Eq. 5a shows) is asymmetric and peaks at a small positive auditory delay (red curve in Fig. 2a and b). This is a widespread characteristic of empirical data (see, e.g., van Eijk et al., 2008, 2010; Vroomen & Keetels, 2010) and is the result of unequal rates λ_v and λ_a. The function is symmetric only when $\lambda_v = \lambda_a$, and, regardless of symmetry or lack thereof, its peak is located at

$$\Delta t_{peak} = \arg\ \max[\Psi_{SJ3\text{-}S}(\Delta t)] = \arg\ \max[\Psi_{SJ2\text{-}S}(\Delta t)] = \delta\frac{\lambda_a - \lambda_v}{\lambda_a + \lambda_v} - \tau \quad (17)$$

(see Section A of the Supplementary Material). In the case illustrated in Fig. 2a and b, the peak occurs at Δt_{peak} = 40 ms.

Second, note also from the shape of the psychometric functions for the SJ3 task in Fig. 2a that $\Psi_{SJ3\text{-}AF}$ (black curve) and $\Psi_{SJ3\text{-}VF}$ (blue curve) both cross $\Psi_{SJ3\text{-}S}$ (red curve) once. These crossings represent the *synchrony boundaries* that render the *synchrony range* as an index of performance in SJ3 tasks. In general, the functions cross *at most* once, and the conditions under which they would not cross are easily determined (see Section A of the Supplementary Material): $\Psi_{SJ3\text{-}AF}$ and $\Psi_{SJ3\text{-}S}$ cross only if $\delta \geq \ln(2)/2\lambda_v$. Given this requirement, if $\exp(-2\delta\lambda_a) \leq (\lambda_v - \lambda_a)/\lambda_v$, the crossing (i.e., the *audio-first synchrony boundary*) occurs at the location $\Delta t_{AFS\text{-}3} \leq -\delta - \tau$ given by

$$\Delta t_{AFS\text{-}3} = \delta - \tau + \frac{1}{\lambda_a}\ln\left[\frac{\lambda_a + \lambda_v}{\lambda_v[2\exp(2\delta\lambda_a) - 1]}\right] \quad (18a)$$

and otherwise the crossing occurs within the interval $(-\delta - \tau, \delta - \tau)$ at the solution of

$$2\lambda_a \exp[-\lambda_v(\delta + \Delta t_{AFS\text{-}3} + \tau)] = \lambda_a + \lambda_v - \lambda_v \exp[-\lambda_a(\delta - \Delta t_{AFS\text{-}3} - \tau)]. \quad (18b)$$

On the other hand, $\Psi_{SJ3\text{-}VF}$ and $\Psi_{SJ3\text{-}S}$ analogously cross only if $\delta \geq \ln(2)/2\lambda_a$. Given this requirement, if $\exp(-2\delta\lambda_v) \leq (\lambda_a - \lambda_v)/\lambda_a$, the crossing (i.e., the *video-first synchrony boundary*) occurs at the location $\Delta t_{VFS\text{-}3} \geq \delta - \tau$ given by

$$\Delta t_{VFS\text{-}3} = \frac{1}{\lambda_v}\ln\left[\frac{\lambda_a[2\exp(2\delta\lambda_v) - 1]}{\lambda_a + \lambda_v}\right] - \delta - \tau \quad (19a)$$

and otherwise the crossing occurs within the interval $(-\delta - \tau, \delta - \tau)$ at the solution of

$$\lambda_a \exp[-\lambda_v(\delta + \Delta t_{VFS\text{-}3} + \tau)] = \lambda_a + \lambda_v - 2\lambda_v \exp[-\lambda_a(\delta - \Delta t_{VFS\text{-}3} - \tau)]. \quad (19b)$$

For the example in Fig. 2a, the AF synchrony boundary occurs at $\Delta t_{AFS\text{-}3}$ = −22.07 ms, and the VF synchrony boundary occurs at $\Delta t_{VFS\text{-}3}$ = 93.06 ms. The midpoint $\Delta t_{SJ3\text{-}PSS}$ of these boundaries is the PSS in SJ3 tasks. In this example, $\Delta t_{SJ3\text{-}PSS} = (\Delta t_{AFS\text{-}3} + \Delta t_{VFS\text{-}3})/2$ = 35.49 ms, and, owing to the asymmetry implied when $\lambda_v \neq \lambda_a$, it does not coincide with the peak of $\Psi_{SJ3\text{-}S}$ (Δt_{peak} = 40 ms). Also, the *synchrony range* (i.e., the width of the region delimited by these boundaries) is $\Delta t_{VFS\text{-}3} - \Delta t_{AFS\text{-}3}$ = 115.13 ms.

Synchrony boundaries in SJ2 tasks are defined at the crossings of $\Psi_{SJ2\text{-}S}$ and $\Psi_{SJ2\text{-}A}$, which occur when the functions evaluate to 1/2 (see Fig. 2b). The functions cross only if $\delta \geq (\lambda_a + \lambda_v)\ln(2)/2\lambda_a\lambda_v$ (see Section A of the Supplementary Material). When this requirement holds, if $\exp(-2\delta\lambda_a) \leq (\lambda_v - \lambda_a)/2\lambda_v$, the AF synchrony boundary occurs at the location $\Delta t_{AFS\text{-}2} \leq -\delta - \tau$ given by

$$\Delta t_{AFS\text{-}2} = \delta - \tau + \frac{1}{\lambda_a}\ln\left[\frac{\lambda_a + \lambda_v}{2\lambda_v[\exp(2\delta\lambda_a) - 1]}\right] \quad (20a)$$

and otherwise it occurs within the interval $(-\delta - \tau, \Delta t_{peak}]$ at the solution of

$$\lambda_a \exp[-\lambda_v(\delta + \Delta t_{AFS\text{-}2} + \tau)] + \lambda_v \exp[-\lambda_a(\delta - \Delta t_{AFS\text{-}2} - \tau)] = (\lambda_a + \lambda_v)/2. \quad (20b)$$

On the other hand, if $\exp(-2\delta\lambda_v) \leq (\lambda_a - \lambda_v)/2\lambda_a$, the VF synchrony boundary occurs at the location $\Delta t_{VFS\text{-}2} \geq \delta - \tau$ given by

$$\Delta t_{\text{VFS-2}} = \frac{1}{\lambda_v}\ln\left[\frac{2\lambda_a[\exp(2\delta\lambda_v)-1]}{\lambda_a+\lambda_v}\right] - \delta - \tau \quad (21a)$$

and otherwise it occurs within the interval $[\Delta t_{\text{peak}}, \delta - \tau)$ at the solution of

$$\lambda_a \exp[-\lambda_v(\delta + \Delta t_{\text{VFS-2}} + \tau)] + \lambda_v \exp[-\lambda_a(\delta - \Delta t_{\text{VFS-2}} - \tau)] = (\lambda_a + \lambda_v)/2. \quad (21b)$$

Note that Eq. 21b has the same form as Eq. 20b, which may have two solutions in the interval $(-\delta - \tau, \delta - \tau)$.

For the example in Fig. 2b, boundaries exist because $\delta = 60 \geq (\lambda_a + \lambda_v)\ln(2)/2\lambda_a\lambda_v = 31.192$. Then, the AF synchrony boundary occurs at $\Delta t_{\text{AFS-2}} = -21.36$ ms, and the VF synchrony boundary occurs at $\Delta t_{\text{VFS-2}} = 88.54$ ms. The midpoint $\Delta t_{\text{SJ2-PSS}}$ of these two boundaries is the PSS in SJ2 tasks, which amounts to $\Delta t_{\text{SJ2-PSS}} = (\Delta t_{\text{AFS-2}} + \Delta t_{\text{VFS-2}})/2 = 33.59$ ms, and, again owing to the asymmetry implied when $\lambda_v \neq \lambda_a$, it does not coincide with the peak of $\Psi_{\text{SJ2-S}}$ ($\Delta t_{\text{peak}} = 40$ ms). The synchrony range is $\Delta t_{\text{VFS-2}} - \Delta t_{\text{AFS-2}} = 109.90$ ms. Note that these indices differ slightly from those computed above for the SJ3 task because of the different criteria used to define landmarks on $\Psi_{\text{SJ3-S}}$ and $\Psi_{\text{SJ2-S}}$.

Finally, note in Fig. 2c that the shape of $\Psi_{\text{TOJ-VF}}$ varies with the response bias parameter ξ. As seen in Eq. 6b, $\xi = 0$ makes $\Psi_{\text{TOJ-VF}}$ identical to $\Psi_{\text{SJ3-VF}}$ (blue curve in Fig. 2c, which is identical to the blue curve in Fig. 2a), whereas $\xi = 1$ makes $\Psi_{\text{TOJ-VF}}$ identical to $1 - \Psi_{\text{SJ3-AF}}$ (black curve in Fig. 2c, which is one minus the black curve in Fig. 2a). Intermediate values for ξ define a transition between these two curves, as illustrated for $\xi \in \{.2, .5, .8\}$ in Fig. 2c (gray curves). The obvious consequence is that the PSS (i.e., the value $\Delta t_{\text{TOJ-PSS}}$ satisfying $\Psi_{\text{TOJ-VF}}(\Delta t_{\text{TOJ-PSS}})=.5$) is affected by the response bias of the observer and can be anywhere between the corresponding values in $\Psi_{\text{SJ3-AF}}$ and $\Psi_{\text{SJ3-VF}}$. The location of the PSS in $\Psi_{\text{TOJ-VF}}$ can also be easily determined (see Section A of the Supplementary Material). It is nevertheless more useful (as will be seen below) to determine the general location of the point $\Delta t_{\text{TOJ-}p}$ at which $\Psi_{\text{TOJ-VF}}(\Delta t_{\text{TOJ-}p}) = p$ (for arbitrary $0 < p < 1$), from which $\Delta t_{\text{TOJ-PSS}} \equiv \Delta t_{\text{TOJ-.5}}$. Thus, if $\lambda_v(1-\xi)\exp(-2\delta\lambda_a) \geq \lambda_a p + \lambda_v(p-\xi)$ then

$$\Delta t_{\text{TOJ-}p} = \frac{1}{\lambda_a}\ln\left[\frac{p(\lambda_a+\lambda_v)}{\lambda_v(1-\xi+\xi\exp[2\delta\lambda_a])}\right] + \delta - \tau; \quad (22a)$$

else, if $\xi\lambda_a \exp(-2\delta\lambda_v) \geq \lambda_v(1-p) - \lambda_a(p-\xi)$, then

$$\Delta t_{\text{TOJ-}p} = \frac{1}{\lambda_v}\ln\left[\frac{\lambda_a[\xi+(1-\xi)\exp(2\delta\lambda_v)]}{(1-p)(\lambda_a+\lambda_v)}\right] - \delta - \tau; \quad (22b)$$

and, otherwise, $\Delta t_{\text{TOJ-}p}$ is at the point within $(-\delta - \tau, \delta - \tau)$ satisfying

$$(1-\xi)\lambda_v \exp\left[-\lambda_a\left(\delta - \Delta_{\text{TOJ-}p} - \tau\right)\right] - \xi\lambda_a \exp\left[-\lambda_v\left(\delta + \Delta_{\text{TOJ-}p} + \tau\right)\right] = (\lambda_a + \lambda_v)(p - \xi). \quad (22c)$$

For the examples in Fig. 2c, the PSS is $\Delta t_{\text{TOJ-PSS}} = 97.26$ ms when $\xi = 0$ (blue curve), $\Delta t_{\text{TOJ-PSS}} = 85.87$ ms when $\xi = .2$ (solid gray curve), $\Delta t_{\text{TOJ-PSS}} = 53.86$ ms when $\xi = .5$ (long-dashed gray curve), $\Delta t_{\text{TOJ-PSS}} = -6.23$ ms when $\xi = .8$ (short-dashed gray curve), and $\Delta t_{\text{TOJ-PSS}} = -22.74$ ms when $\xi = 1$ (black curve). As was discussed above, the cases $\xi = 0$ and $\xi = 1$ refer to the 50 % point $\Delta t_{\text{VF-PSS}}$ on $\Psi_{\text{SJ3-VF}}$ and the 50 % point $\Delta t_{\text{AF-PSS}}$ on $\Psi_{\text{SJ3-AF}}$, respectively.

Therefore, even if the underlying parameters are invariant across tasks, PSSs in SJ2 or SJ3 tasks differ minimally and only as a result of the different landmarks used to define the PSS from the otherwise identical functions $\Psi_{\text{SJ2-S}}$ and $\Psi_{\text{SJ3-S}}$. In the preceding example, these are $\Delta t_{\text{SJ2-PSS}} = 33.59$ ms and $\Delta t_{\text{SJ3-PSS}} = 35.49$ ms. In contrast, and owing to the participation of the response bias parameter ξ in TOJ tasks, PSSs in these tasks may vary greatly. In the preceding example, they may range from $\Delta t_{\text{TOJ-PSS}} = -22.74$ ms when $\xi = 1$ to $\Delta t_{\text{TOJ-PSS}} = 97.26$ ms when $\xi = 0$. Note that even for an unbiased observer (i.e., $\xi = .5$), the PSS will be $\Delta t_{\text{TOJ-PSS}} = 53.86$ ms in the preceding example, which is meaningfully different from its counterparts in SJ2 or SJ3 tasks (33.59 and 35.49 ms, respectively). The latter characteristic is due to the asymmetry caused by $\lambda_v \neq \lambda_a$; when $\lambda_v = \lambda_a$ instead, the PSS in a TOJ task with $\xi = .5$ matches the PSS in SJ2 and SJ3 tasks.

These features of the model are consistent with the empirically observed similarity of PSS estimates from SJ2 and SJ3 tasks and the discrepant estimates from TOJ tasks, and also with the lack of correlation with estimates from SJ2 and SJ3 tasks. In other words, the empirical "observation that the TOJ 50 % points occur virtually anywhere in the synchrony range" (van Eijk et al., 2008, p. 964) may have its origin in the response bias parameter ξ, which plays a major role in shaping $\Psi_{\text{TOJ-PSS}}$ even if the remaining parameters have the same values as in SJ2 and SJ3 tasks.

Appendix 2 Tables with parameter estimates and descriptive statistics

Table 1 Estimated parameters for each task in the flash–click experiment. Dashes indicate parameters not included in the reduced model that was selected for the corresponding observer and task. Stars on the right of the block for each task indicate that the model was rejected by a .05-size chi-square test with the degrees of freedom indicated in the header row for each task, where k is the number of parameters in the model for each particular observer

	SJ2 task (15 - k degrees of freedom)								SJ3 task (30 - k degrees of freedom)										TOJ task (15 - k degrees of freedom)							
Obs.	$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}	$\kappa_{AF\text{-}S}$	$\kappa_{S\text{-}AF}$	$\kappa_{VF\text{-}AF}$		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ξ	ε_{VF}
1	5.00	69.44	10.64	101.35	–	–	–		36.41	47.52	-31.47	106.05	–	–	.095	–	–	0.44		47.94	39.43	-51.93	56.12	–	.475	–
2	53.89	66.20	-11.92	159.00	–	–	–		60.63	30.75	-26.43	122.93	.034	–	.128	0.39	–	0.57		137.14	159.08	-6.56	60.73	–	.482	–
3	27.92	19.19	-69.06	127.21	–	–	.072	*	27.06	25.31	-45.24	108.29	.013	–	.153	0.00	–	0.23	*	5.00	87.22	115.97	66.76	.008	.122	–
4	66.50	47.64	-52.60	94.16	–	–	–		56.68	49.98	-45.28	96.14	.014	–	.060	0.00	–	1.00	*	22.61	79.03	-55.82	69.63	.005	.826	–
5[a]	–	–	–	–	–	–	–		–	–	–	–	–	–	–	–	–	–		–	–	–	–	–	–	–
6	36.65	5.00	-33.21	81.51	–	.418	.010		5.00	19.38	1.81	53.85	.024	.636	.011	0.00	0.24	1.00		35.53	51.74	131.68	114.85	–	.043	.020
7	36.79	25.36	-27.90	84.48	–	–	–		28.66	42.36	-11.49	108.72	.023	–	.009	0.76	–	1.00		5.00	50.24	-22.80	62.93	.011	.470	–
8	39.58	29.51	-12.76	125.41	–	–	.055		28.21	35.48	-4.91	123.35	.065	–	.045	0.21	–	0.85		50.29	26.13	-62.67	0.00	.032	.331	.016
9	54.79	58.35	-41.97	101.58	–	–	.064		40.64	46.37	-23.34	109.24	–	.187	.024	–	0.57	1.00		5.00	156.66	21.77	123.88	.056	.777	–
10	60.00	28.46	-73.61	177.06	–	–	–		37.13	23.96	-52.14	150.55	.028	.112	.043	0.00	0.24	1.00	*	18.40	77.46	-67.12	138.06	.013	.933	–
11	53.02	61.39	68.98	166.44	–	–	.045		49.10	61.35	69.20	160.29	.227	–	.022	0.17	–	1.00		5.00	61.57	125.55	117.75	.090	.245	–
12	37.83	51.66	-24.70	161.65	.071	–	.123		33.97	50.22	10.08	172.60	.076	.022	–	0.75	0.60	–		74.63	36.18	-62.59	124.82	–	.819	–
mean[b]	42.47	41.58	-24.37	125.44	.071	.418	.062		36.25	39.33	-14.47	119.27	.056	.239	.059	0.26	0.41	0.79		35.21	74.98	5.95	85.05	.031	.502	.018
SD[b]	17.51	21.30	37.96	33.74	.000	.000	.034		15.69	12.66	32.74	31.42	.064	.237	.048	0.30	0.17	0.32		40.09	43.07	77.08	39.98	.029	.289	.002

[a] Observer 5 did not participate in the flash–click experiment

[b] Boundary values of 5.00 for $1/\lambda_a$ and $1/\lambda_v$ are excluded from these computations (see Appendix 3 for a discussion)

Table 2 Estimated parameters for each task in the bouncing ball experiment. Dashes indicate parameters not included in the reduced model that was selected for the corresponding observer and task. Stars on the right of the block for each task indicate that the model was rejected by a .05-size chi-square test with the degrees of freedom indicated in the header row for each task, where k is the number of parameters in the model for each particular observer

	SJ2 task (15 - k degrees of freedom)								SJ3 task (30 - k degrees of freedom)										TOJ task (15 -k degrees of freedom)							
Obs.	$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}	$\kappa_{AF\text{-}S}$	$\kappa_{S\text{-}AF}$	$\kappa_{VF\text{-}AF}$		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ξ	ε_{VF}
1	29.90	50.51	-57.99	93.64	.007	–	–		21.11	30.75	-51.37	98.15	.017	.128	–	0.00	0.00	–	*	38.33	5.00	-141.55	138.58	.029	.975	.008
2	23.05	41.06	-15.65	108.64	–	–	–		24.47	43.88	-25.65	95.05	.022	–	.040	0.63	–	0.50		66.28	5.00	-108.02	38.89	.022	.330	.008
3	15.84	37.87	-25.55	119.52	.016	–	.055	*	27.21	33.16	-53.75	128.89	.018	–	–	0.00	–	–	*	20.40	37.80	-46.75	56.14	–	.553	–
4	47.82	5.95	-104.47	104.53	–	–	.094		26.06	15.32	-59.97	104.82	.034	.107	.073	0.38	0.13	0.07		43.62	21.47	-66.37	84.51	.019	.851	.027
5	21.93	23.57	-25.99	123.36	–	–	–		25.79	33.42	-11.70	115.49	.055	.078	.068	0.29	0.29	0.45		38.60	5.00	-128.42	132.10	.014	.910	–
6	20.88	15.95	-39.08	67.72	–	–	.009		5.00	14.91	-23.38	71.31	–	.109	.003	–	0.00	1.00		46.74	37.79	-10.97	47.44	–	.430	–
7	20.90	16.07	-60.71	104.27	.003	–	.036		37.71	20.41	-44.88	106.33	–	–	.007	–	–	0.49		5.00	53.34	60.61	52.35	.007	.160	–
8	25.24	25.78	-13.60	115.86	–	–	–		22.51	15.41	-19.34	108.03	.083	.035	.040	0.00	0.00	0.82	*	20.91	48.84	16.91	51.33	.009	.661	–
9	27.69	43.90	-4.60	129.29	.030	.093	–		39.27	29.05	-33.19	102.88	.023	.139	.072	0.30	0.54	0.69		45.49	34.19	-67.93	92.19	.022	.851	.028
10	6.33	39.76	-26.08	123.95	–	–	–		16.13	5.00	-61.60	93.36	.003	.049	.020	1.00	0.62	0.00		6.92	67.33	106.32	56.37	.012	.097	–
11	25.06	45.78	-2.31	151.14	–	–	–		21.08	56.34	29.31	178.59	.188	–	.040	0.83	–	1.00		38.27	5.00	-34.34	78.59	.067	.375	.083
12	5.00	89.60	39.68	129.31	.079	–	–		44.14	43.14	-10.86	113.41	.015	.111	.014	0.00	0.06	1.00		88.11	121.44	28.11	127.62	–	.515	–
mean[a]	20.47	36.32	-28.03	114.27	.027	.093	.049		25.87	28.37	-30.53	109.69	.046	.069	.038	0.46	0.36	0.60		37.82	35.25	-32.70	79.67	.022	.559	.031
SD[a]	10.28	20.91	34.28	20.07	.028	.000	.031		10.18	14.21	25.07	24.72	.052	.048	.025	0.41	0.38	0.35		23.22	34.26	72.99	34.21	.017	.282	.028

[a] Boundary values of 5.00 for $1/\lambda_a$ and $1/\lambda_v$ are excluded from these computations (see Appendix 3 for a discussion)

Table 3 Estimated parameters of the joint model in the flash–click experiment. Dashes indicate parameters not included in the reduced model that was selected for the corresponding observer. Stars indicate that the model was rejected by a .05-size chi-square test with $60 - k$ degrees of freedom, where k is the actual number of parameters in the model for each particular observer

		common parameters			SJ2 task				SJ3 task							TOJ task			
Obs.		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}	δ	ε_{AF}	ε_S	ε_{VF}	$\kappa_{AF\text{-}S}$	$\kappa_{S\text{-}AF}$	$\kappa_{VF\text{-}AF}$	δ	ε_{AF}	ξ	ε_{VF}
1		40.16	46.66	-39.30	99.79	–	–	–	106.23	–	–	.089	–	–	0.47	56.75	–	.488	–
2		54.16	39.82	-21.09	161.96	–	.060	.064	122.60	.049	–	.116	0.55	–	0.62	98.43	.062	.458	.103
3	*	33.10	10.57	-78.06	126.47	–	–	.087	126.08	–	.065	.103	–	0.06	0.00	81.60	.015	.784	.034
4	*	58.85	51.40	-46.95	93.17	–	–	–	96.58	.014	–	.055	0.00	–	1.00	53.48	–	.361	.018
5[a]		–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
6		32.29	10.02	-26.43	80.70	–	.423	.012	57.14	.020	.616	.012	0.00	0.31	1.00	45.81	.020	.605	.030
7		30.88	41.29	-13.53	86.15	–	–	–	107.90	.020	–	.009	0.78	–	1.00	72.27	.008	.203	–
8		36.62	32.34	-11.78	126.32	–	–	.046	124.09	.057	–	.046	0.11	–	0.83	40.12	.039	.085	.014
9	*	42.18	45.30	-30.87	113.96	–	.139	.091	108.49	–	.184	.024	–	0.53	1.00	113.76	.049	.557	.107
10	*	41.76	32.43	-50.19	181.98	–	.034	–	150.47	.026	.110	.045	0.00	0.22	1.00	78.57	.009	.792	.034
11		50.83	61.04	70.06	166.18	–	–	.045	161.72	.218	–	.023	0.12	–	1.00	101.31	.099	.224	–
12	*	41.75	46.62	-14.18	156.84	.052	–	.187	171.68	.089	.021	–	0.78	0.56	–	107.17	.034	.800	–
mean		42.05	37.96	-23.85	126.68	.012	.164	.076	121.18	.062	.199	.048	0.29	0.34	0.72	77.21	.037	.487	.049
SD		8.73	15.15	35.21	33.76	.000	.154	.052	30.68	.064	.215	.037	0.33	0.19	0.32	24.59	.028	.238	.037

[a] Observer 5 did not participate in the flash–click experiment

Table 4 Estimated parameters of the joint model in the bouncing ball experiment. Dashes indicate parameters not included in the reduced model that was selected for the corresponding observer. Stars indicate that the model was rejected by a .05-size chi-square test with $60 - k$ degrees of freedom, where k is the actual number of parameters in the model for each particular observer

		common parameters			SJ2 task				SJ3 task							TOJ task			
Obs.		$1/\lambda_a$	$1/\lambda_v$	τ	δ	ε_{AF}	ε_S	ε_{VF}	δ	ε_{AF}	ε_S	ε_{VF}	$\kappa_{AF\text{-}S}$	$\kappa_{S\text{-}AF}$	$\kappa_{VF\text{-}AF}$	δ	ε_{AF}	ξ	ε_{VF}
1		32.62	21.26	-78.86	99.10	.007	.094	.051	108.39	.015	.166	–	0.00	0.00	–	96.81	.031	.980	.017
2		27.28	39.75	-26.08	109.03	–	–	–	94.51	.019	–	.050	0.60	–	0.44	58.04	.033	.267	–
3	*	19.19	39.39	-33.62	118.59	.017	–	.041	124.71	.019	–	–	0.00	–	–	58.07	–	.467	–
4	*	38.01	21.68	-72.05	97.83	–	–	.128	109.73	.029	.109	.054	0.35	0.10	0.09	93.65	.022	.865	.025
5		24.97	24.06	-25.58	123.32	–	–	–	118.86	.059	.096	.069	0.33	0.31	0.43	66.28	.018	.859	.022
6	*	16.92	18.11	-31.81	69.53	–	–	.008	67.72	–	.105	.003	–	0.00	1.00	65.62	.016	.628	–
7	*	36.47	13.60	-63.04	102.92	–	–	.053	109.75	–	–	.008	–	–	0.46	44.82	.008	.766	.013
8		25.86	22.80	-16.72	116.61	–	–	–	103.62	.096	–	.036	0.00	–	0.90	66.12	.008	.663	.004
9		35.33	39.02	-18.51	127.49	.028	.085	–	101.71	.024	.127	.069	0.32	0.51	0.73	62.29	.025	.782	.040
10	*	16.01	32.88	-42.97	122.42	–	–	–	91.72	.003	.035	–	1.00	0.53	–	78.49	.023	.891	.013
11	*	29.65	44.79	-2.83	151.30	–	–	–	171.21	.239	–	.042	0.87	–	1.00	76.63	.069	.427	.073
12	*	54.95	47.14	-21.36	124.81	.057	–	.071	119.47	–	.153	–	–	0.18	–	129.50	.050	.547	.052
mean		29.77	30.37	-36.12	113.58	.027	.090	.059	110.12	.056	.113	.041	0.38	0.23	0.63	74.69	.027	.678	.029
SD		10.39	10.90	22.62	19.33	.019	.004	.036	23.46	.070	.040	.023	0.35	0.20	0.31	21.79	.017	.208	.021

Appendix 3

In the model, $1/\lambda_a$ and $1/\lambda_v$ represent the standard deviation of the arrival latencies (in milliseconds) of auditory and visual signals, respectively, at the central mechanism, and it is unrealistic to assume that these standard deviations can be arbitrarily small. This is the reason that their lower bounds in the parameter space for the optimization algorithm were set at 5 ms. As shown here, estimates of any of these two parameters at the boundary value reveals only that the data are inconclusive about the value of that particular parameter. Consider the case of the separate SJ2 task for observer 6 in the flash–click experiment (data and model curve plotted in the second panel down the left column of Fig. 5 and reproduced also in the top panel of Fig. 15). From Table 1 in Appendix 2, parameter estimates are ($1/\lambda_a$, $1/\lambda_v$, τ, δ, ε_S, ε_{VF}) = (36.65, 5.00, −33.21, 81.51, 0.418, 0.010), and the top panel of Fig. 15 indeed shows that the model curve describes the data very accurately despite the boundary estimate. The consequences of removing the boundary on estimates of $1/\lambda_v$ are illustrated for this particular case in the bottom panel of Fig. 15: The model curve only has sharper corners (indicated by arrows; compare with the smoother corners in the top panel) but with no actual improvement in how the curve accommodates the data. The nominal fit is better when the boundary on $1/\lambda_v$ is removed,

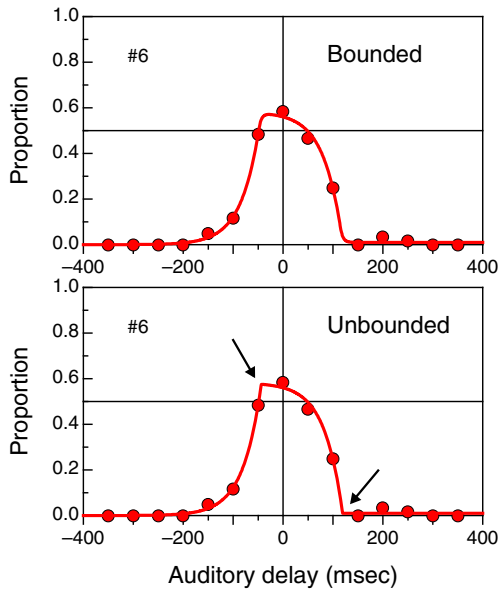

Fig. 15 Inconclusive data about the value of parameter λ_v, which renders a boundary estimate for this parameter and the model curve in the top panel (replotted from our Fig. 5 at a larger size). Removal of the boundary on this parameter renders the result in the bottom panel and only produces sharper corners (arrows) without any meaningful change in the shape of the fitted curve or the estimated values for the remaining parameters. The cost of removing the boundary is that the estimated standard deviation of visual arrival latencies is unrealistically purported to be virtually zero

but, in terms of parameter estimates, those obtained in the unbounded case are ($1/\lambda_a$, $1/\lambda_v$, τ, δ, ε_S, ε_{VF}) = (36.66, 0.0002, −37.92, 81.53, 0.418, 0.010). In other words, no meaningful or relevant change (if a change at all) occurs in parameter estimates other than $1/\lambda_v$, which now turns out to be 0.0002. But visual arrival latencies cannot be accepted to have this standard deviation, and this is the reason for our setting an arbitrary lower bound at 5 ms.

Similar outcomes were observed in all the other cases listed in Tables 1 and 2 in which estimates of $1/\lambda_a$ or $1/\lambda_v$ attained boundary values. The common characteristic of all these cases is that the fixed set of auditory delays that was used with all observers turned out to sample insufficiently the psychometric functions for some of them: In the case shown in Fig. 15, all auditory delays at or below −200 ms and at or above 150 ms render uninformative data, and only the six auditory delays between −150 and 100 ms are useful to estimate model parameters. A way around the uncertainty as to which set of auditory delays would be needed for each observer is to use adaptive sampling methods analogous to those that efficiently solve this problem in other contexts (García-Pérez & Alcalá-Quintana, 2005). Note, however, that the example considered here reveals that all the remaining model parameters are reasonably well estimated, and the same was true in the remaining cases affected by boundary estimates. A reasonable replacement for the boundary estimate in these cases could, perhaps, be sought post hoc by increasing the value of the affected parameter until the model curve begins to part company with the data, but this is a rather ad hoc method with uncertain outcomes. We decided instead to regard boundary estimates of $1/\lambda_a$ and $1/\lambda_v$ as missing values in all statistical analyses.

References

Alcalá-Quintana, R., & García-Pérez, M. A. (2004). The role of parametric assumptions in adaptive Bayesian estimation. *Psychological Methods, 9,* 250–271. doi:10.1037/1082-989X.9.2.250

Alcalá-Quintana, R., & García-Pérez, M. A. (2011). A model for the time-order error in contrast discrimination. *Quarterly Journal of Experimental Psychology, 64,* 1221–1248. doi:10.1080/17470218.2010.540018

Alcalá-Quintana, R., & García-Pérez, M. A. (2012). Fitting psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines. Manuscript submitted for publication.

Allan, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception & Psychophysics, 18,* 29–36. doi:10.3758/BF03199363

Allan, L. G., & Kristofferson, A. B. (1974). Successiveness discrimination: Two models. *Perception & Psychophysics, 15,* 37–46. doi:10.3758/BF03205825

Ben-Yehudah, G., Sackett, E., Malchi-Ginzberg, L., & Ahissar, M. (2001). Impaired temporal contrast sensitivity in dyslexics is specific to retain-and-compare paradigms. *Brain, 124,* 1381–1395. doi:10.1093/brain/124.7.1381

Colonius, H., & Diederich, A. (2011). Computing an optimal time window of audiovisual integration in focused attention tasks: Illustrated by studies on effect of age and prior knowledge. *Experimental Brain Research, 212,* 327–337. doi:10.1007/s00221-011-2732-x

Diederich, A., & Colonius, H. (2011). Modeling multisensory processes in saccadic responses: Time-window-of-integration model. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 253–276). Boca Raton: CRC Press.

Donohue, S. E., Woldorff, M. G., & Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Attention, Perception, & Psychophysics, 72,* 1120–1129. doi:10.3758/APP.72.4.1120

Fujisaki, W., & Nishida, S. (2009). Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. *Experimental Brain Research, 198,* 245–259. doi:10.1007/s00221-009-1870-x

García-Pérez, M. A. (1994). Parameter estimation and goodness-of-fit testing in multinomial models. *British Journal of Mathematical and Statistical Psychology, 47,* 247–282. doi:10.1111/j.2044-8317.1994.tb01037.x

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties. *Vision Research, 38,* 1861–1881. doi:10.1016/S0042-6989(97)00340-4

García-Pérez, M. A. (2010). Denoising forced-choice detection data. *British Journal of Mathematical and Statistical Psychology, 63,* 75–100. doi:10.1348/000711009X424057

García-Pérez, M. A., & Alcalá-Quintana, R. (2005). Sampling plans for fitting the psychometric function. *Spanish Journal of Psychology*, *8*, 256–289. Retrieved from http://www.ucm.es/info/Psi/docs/journal/v8_n2_2005/art256.pdf

García-Pérez, M. A., & Alcalá-Quintana, R. (2010a). The difference model with guessing explains interval bias in two-alternative forced-choice detection procedures. *Journal of Sensory Studies, 25,* 876–898. doi:10.1111/j.1745-459X.2010.00310.x

García-Pérez, M. A., & Alcalá-Quintana, R. (2010b). Reminder and 2AFC tasks provide similar estimates of the difference limen: A reanalysis of data from Lapid, Ulrich, and Rammsayer (2008) and a discussion of Ulrich and Vorberg (2009). *Attention, Perception, & Psychophysics, 72,* 1155–1178. doi:10.3758/APP.72.4.1155

García-Pérez, M. A., & Alcalá-Quintana, R. (2011a). Improving the estimation of psychometric functions in 2AFC discrimination tasks. *Frontiers in Psychology: Quantitative Psychology and Measurement, 2,* 96. doi:10.3389/fpsyg.2011.00096

García-Pérez, M. A., & Alcalá-Quintana, R. (2011b). Interval bias in 2AFC detection tasks: Sorting out the artifacts. *Attention, Perception, & Psychophysics, 73,* 2332–2352. doi:10.3758/s13414-011-0167-x

García-Pérez, M. A., & Alcalá-Quintana, R. (2012a). Response errors explain the failure of independent-channels models of perception of temporal order. *Frontiers in Psychology: Perception Science, 3,* 94. doi:10.3389/fpsyg.2012.00094

García-Pérez, M. A., & Alcalá-Quintana, R. (2012b). Shifts of the psychometric function: Distinguishing bias from perceptual effects. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2012.708761

García-Pérez, M. A., Alcalá-Quintana, R., Woods, R. L., & Peli, E. (2011). Psychometric functions for detection and discrimination with and without flankers. *Attention, Perception, & Psychophysics, 73,* 829–853. doi:10.3758/s13414-010-0080-8

García-Pérez, M. A., & Núñez-Antón, V. (2001). Small-sample comparisons for power-divergence goodness-of-fit statistics for symmetric and skewed simple null hypotheses. *Journal of Applied Statistics, 28,* 855–874. doi:10.1080/02664760120074942

García-Pérez, M. A., & Núñez-Antón, V. (2004). Small-sample comparisons for goodness-of-fit statistics in one-way multinomials with composite hypotheses. *Journal of Applied Statistics, 31,* 161–181. doi:10.1080/0266476032000148849

Gilchrist, J. M., Jerwood, D., & Ismaiel, H. S. (2005). Comparing and unifying slope estimates across psychometric function models. *Perception & Psychophysics, 67,* 1289–1303. doi:10.3758/BF03193560

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research, 186,* 517–524. doi:10.1007/s00221-007-1253-0

Heath, R. A. (1984). Response time and temporal order judgement in vision. *Australian Journal of Psychology, 36,* 21–34. doi:10.1080/00049538408255075

Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62,* 423–432. doi:10.1037/h0045283

Holm, H., & Alouini, M.-S. (2004). Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution. *IEEE Transactions on Communications, 52,* 1367–1376. doi:10.1109/TCOMM.2004.833019

Jaśkowski, P. (1991a). Two-stage model for order discrimination. *Perception & Psychophysics, 50,* 76–82. doi:10.3758/BF03212206

Jaśkowski, P. (1991b). Perceived onset simultaneity of stimuli with unequal durations. *Perception, 20,* 715–726. doi:10.1068/p200715

Jaśkowski, P. (1993). Selective attention and temporal-order judgment. *Perception, 22,* 681–689. doi:10.1068/p220681

Kristofferson, A. B., & Allan, L. G. (1973). Successiveness and duration discrimination. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 737–749). New York: Academic Press.

Küchler, U., & Tappe, S. (2008). On the shapes of bilateral Gamma densities. *Statistics and Probability Letters, 78,* 2478–2484. doi:10.1016/j.spl.2008.02.039

Landerl, K., & Willburger, E. (2010). Temporal processing, attention, and learning disorders. *Learning and Individual Differences, 20,* 393–401. doi:10.1016/j.lindif.2010.03.008

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45,* 255–268. doi:10.2307/2532051

Mitrani, L., Shekerdjiiski, S., & Yakimoff, N. (1986). Mechanisms and asymmetries in visual perception of simultaneity and temporal order. *Biological Cybernetics, 54,* 159–165. doi:10.1007/BF00356854

Nicholls, M. E. R., Lew, M., Loetscher, T., & Yates, M. J. (2011). The importance of response type to the relationship between temporal order and numerical magnitude. *Attention, Perception, & Psychophysics, 73,* 1604–1613. doi:10.3758/s13414-011-0114-x

Numerical Algorithms Group. (1999). *NAG Fortran Library Manual, Mark 19*. Oxford: Author.

Occelli, V., Spence, C., & Zampini, M. (2011). Audiotactile interactions in temporal perception. *Psychonomic Bulletin & Review, 18,* 429–454. doi:10.3758/s13423-011-0070-4

Peli, E., & García-Pérez, M. A. (1997). Contrast sensitivity in dyslexia: Deficit or artifact? *Optometry and Vision Science, 74,* 986–988. doi:10.1097/00006324-199712000-00017

Ram-Tsur, R., Faust, M., & Zivotofsky, A. Z. (2006). Sequential processing deficits of reading disabled persons is independent of inter-stimulus interval. *Vision Research, 46,* 3949–3960. doi:10.1016/j.visres.2006.07.001

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology, 47,* 333–366. doi:10.1016/S0010-0285(03)00035-5

Shore, D. I., Spry, E., & Spence, C. (2002). Confusing the mind by crossing the hands. *Cognitive Brain Research, 14,* 153–163. doi:10.1016/S0926-6410(02)00070-8

Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics, 65,* 318–328. doi:10.3758/BF03194803

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition, 19,* 364–379. doi:10.1016/j.concog.2009.12.001

Stelmach, L. B., & Herdman, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 539–550. doi:10.1037/0096-1523.17.2.539

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., . . . Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B*, *268*, 31–38. doi:10.1098/rspb.2000.1326

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629–685). New York: Academic Press.

Townsend, J. T., & Colonius, H. (2005). Variability of the MAX and MIN statistic: A theory of the quantile spread as a function of sample size. *Psychometrika, 70,* 759–772. doi:10.1007/s11336-001-0916-1

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception & Psychophysics, 42,* 224–239. doi:10.3758/BF03203074

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Perception & Psychophysics, 70,* 955–968. doi:10.3758/PP.70.6.955

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F., & van de Par, S. (2010). Temporal order judgment criteria are affected by synchrony judgment sensitivity. *Attention, Perception, & Psychophysics, 72,* 2227–2235. doi:10.3758/APP.72.8.2227

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research, 185,* 521–529. doi:10.1007/s00221-007-1168-9

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72,* 871–874. doi:10.3758/APP.72.4.871

Yamamoto, S., & Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience, 4,* 759–765. doi:10.1038/89559

Yarrow, K., Jahn, N., Durant, S., & Arnold, D. H. (2011). Shifts of criteria or neural timing? The assumptions underlying timing perception studies. *Consciousness and Cognition, 20,* 1518–1531. doi:10.1016/j.concog.2011.07.003

Yates, M. J., & Nicholls, M. E. R. (2011). Somatosensory prior entry assessed with temporal order judgments and simultaneity judgments. *Attention, Perception, & Psychophysics, 73,* 1586–1603. doi:10.3758/s13414-011-0117-7

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio–visual simultaneity judgments. *Perception & Psychophysics, 67,* 531–544. doi:10.3758/BF03193329

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research, 152,* 198–210. doi:10.1007/s00221-003-1536-z