Visual intensity-dependent response latencies predict perceived audio-visual simultaneity

Ryan Horsfall[I]. Sophie Wuerger[I]. Georg Meyer[I].

[I]University of Liverpool, School of Psychology, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA.

Corresponding author: Ryan Horsfall

Present address: Zochonis Building, The University of Manchester, Manchester, M13 9GB.

Abstract

To form a coherent presentation of the world, the brain needs to accurately combine multiple sensory modalities together in the temporal domain. Judgements on the relative timing of audio-visual stimuli are complex, due to the differing propagation speeds of light and sound through the environment and the nervous system, and the dependence of processing latencies on stimulus intensity (Piéron, 1913). Simultaneity judgement (SJ) and temporal order judgement (TOJ) tasks are often used to assess the temporal mechanisms underlying this binding process. However, these tasks consistently produce measures of perceived simultaneity that are uncorrelated with each other, leading to the suggestion that SJ and TOJ tasks could depend on separate neural mechanisms.

Parise and Ernst's (2016) multisensory correlation detector (MCD) model predicts this lack of correlation by assuming two internal processing stages, a lag computation and a correlation. Here we include and empirically evaluate an intensity-dependent processing delay in the MCD model.

We estimate the points of subjective simultaneity (PSSs) using both SJ and TOJ tasks for four different visual intensities and a fixed auditory sound level. Evaluation of four variants of the intensity-dependent MCD model show that the introduction of an early processing delay can predict the different PSS values obtained in the two respective tasks, without the need for later intensity-dependent multisensorial processing stages. Crucially, this early processing delay can be estimated from simple reaction times.

1. Introduction

A vital function of the human perceptual system is to combine stimuli of different modalities to gain a holistic view of the environment. Sensory signals originating from the same physical event must be associated, while signals caused by unrelated physical events must be segregated. Temporal coincidence is a useful heuristic for these perceptual decisions (Burr, Silva, Cicchini, Banks, & Morrone, 2009; Diederich & Colonius, 2015; Leone & McCourt, 2012, 2013; Meredith, Nemitz, & Stein, 1987). Furthermore, temporally coincident audio-visual stimuli are processed faster (Leone & Mccourt, 2012; 2013; 2015; Miller, 1986; Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012) and are more accurately detected (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Fiebelkorn, Foxe, Butler, & Molholm, 2011) than stimuli which occur separately in time.

However, the differences in propagation speed of the two modalities in the nervous system (King & Palmer, 1985; Raij et al., 2010) and through the environment, mean that two co-occurring signals may reach the respective sensory cortices at differing latencies (Raij et al., 2010). The temporal relationship between stimuli is further compounded by the level-dependent processing speeds of auditory and visual signals. Increasing stimulus intensity leads to reduced action potential latencies (Carrillo-de-la-Peña et al., 1999) and shorter reaction times for visual (Pins & Bonnet, 1996), auditory (Jaśkowski, Rybarczyk, Jaroszyk, & Lemanski, 1995; Ulrich, Rinkenauer, & Miller, 1998) and bimodal audio-visual stimuli (e.g. Harrison, Wuerger, & Meyer, 2010).

1.1. Perceived audio-visual simultaneity for TOJ and SJ tasks

To assess the perceived timing of auditory-visual stimuli, two tasks are commonly used: a simultaneity judgement (SJ) [1] and a temporal order judgement (TOJ) task (e.g. Zampini, Guest, Shore, & Spence, 2005; Zampini, Shore, & Spence, 2003). These require participants to respond to pairs of

[1] Abbreviations: Simultaneity Judgement (SJ). Temporal order judgement (TOJ). Stimulus onset asynchrony (SOA). Point of subjective simultaneity (PSS). Reaction time (RT). Multisensory Correlation Detector (MCD). Corrected Akaike information criterion (AIC_c).

stimuli with different stimulus onset asynchronies (SOA). In the SJ task, participants decide whether the two stimuli occurred at the same time or separately. The point of subjective simultaneity (PSS) is the SOA at which the observers are mostly likely to respond 'simultaneous' (Barnett-Cowan & Harris, 2009; Leone & Mccourt, 2015). In the TOJ task, participants judge which of the two stimuli, auditory or visual, occurred first and the PSS corresponds to the SOA at which the observers respond equally often 'auditory first' and 'visual first' (Leone & Mccourt, 2015). The lack of correlation between the PSS obtained in the two tasks (Love, Petrini, Cheng, & Pollick, 2013; Van Eijk, Kohlrausch, Juola, & Van De Par, 2008) has led to the suggestion that fundamentally different processes underlie these two judgements (e.g. Shore, Spry, & Spence, 2002; Vatakis, Navarra, Soto-Faraco, & Spence, 2008). One idea is the two-stage model of order discrimination (Jaśkowski, 1991), which proposes that two separate mechanisms independently assess the difference in arrival latency of two signals. First simultaneity is evaluated, but if the interval between the two signal arrival times is too large then the relative order is established at an order judgement. There is some support from brain imaging for additional neural activation for TOJs over SJs (Binder, 2015; Miyazaki et al., 2016). Other explanations for the lack of correlation between the TOJ and SJ tasks evoke shifts in decision criteria (Ulrich, Allan, Giray, Schmid, & Vorberg, 1987; Yarrow, Jahn, Durant, & Arnold, 2011).

1.2. Intensity dependence of perceived simultaneity

For visual and auditory stimuli to be perceived as simultaneous, dimmer visual stimuli must be presented earlier than brighter stimuli, relative to an auditory stimulus (e.g. Roufs, 1963). This holds for both the TOJ (Boenke, Deliano, & Ohl, 2009; Leone & Mccourt, 2015) and SJ task (Leone & McCourt, 2015), whereby reducing visual intensity leads to a PSS shift towards the visual leading offsets. Whether this intensity-dependence of perceived simultaneity is a major factor in real-world scenarios is debatable, but it is a robust finding in lab-based conditions. Various mechanisms have been proposed to compensate for these non-temporal stimulus features biasing the perception of

simultaneity (Alais & Carlile, 2005; Kopinska & Harris, 2004; Noel, De Niear, Van der Burg, & Wallace, 2016; Simon, Noel, & Wallace, 2017; Van der Burg, Alais, & Cass, 2013) .

The same intensity manipulation has been shown to have contrasting effects on audio-visual RTs. The intensity dependent latency shifts, observed as a PSS shift in the SJ and TOJ tasks, were not reflected in race model violations (Miller, 1982), i.e. the speeded response to bimodal stimuli, - indicative of multisensory interactions - occurred exclusively at an SOA of 0ms, regardless of visual intensity (Leone & McCourt, 2015). This provides evidence for a distinction between the effect of intensity on perceived simultaneity, and on *bimodal* RTs, whereby intensity dependent differences in processing latencies can be compensated for at the neuronal level (Leone & McCourt, 2013; 2015). Additionally, the effect of intensity on the PSS has been reported to be considerably less than the corresponding effect on unimodal reaction times (Menendez & Lit, 1983; Jaśkowski, 1992), implying that PSS shifts induced by varying relative stimulus intensity may not be driven solely by intensity dependent variation in perceptual latencies. This disparity may be also be caused by methodological differences in stimulus presentation (Menendez & Lit, 1983; Jaśkowski, 1992) and subjective task difficulty (Love, Petrini, Cheng & Pollick, 2013). Furthermore, SJs, TOJs and RTs all involve task-specific decisional criteria (Yarrow et al., 2011; Linares & Holcombe, 2014) which could be a further source of the reported disparity in the effect of intensity on the three tasks.

1.3. Purpose of the current study

Parise and Ernst (2016) proposed a multisensory correlation detector (MCD), in which the observed lack of correlation between the PSSs obtained in the SOJ and TOJ tasks is accounted for by two separate internal computations which are weighted differently in the two tasks, rather than postulating divergent early processing mechanisms (Figure 1). The purpose of the current study is (1) to include an intensity-dependent early component in the MCD model, which transforms the intensity variations into a processing delay. (2) This early intensity-dependent component will be

estimated from reaction time data. (3) We will then test the modified MCD model by comparing it with psychometric functions obtained in a SJ and TOJ task at different intensity levels.

2. A modified Multisensory Correlation Detector

2.1. Model Overview

The original MCD model (Parise & Ernst, 2016) is based upon the Hassenstein-Reichardt detector model of motion perception (Hassenstein & Reichardt, 1958). In the MCD model the perception of simultaneity and temporal order both rely on two detectors which compare the signals of two subunits. The *correlation detector* computes the temporal overlap of the two signals, whereas the *lag detector* outputs the temporal difference. The lack of correlation between the PSSs of SJs and TOJs is explained by a difference in weighting applied to the outputs of the two detectors for the two tasks.

The MCD assumes that both visual and auditory stimuli pass through modality specific low-pass filters (denoted 'Temporal Filtering' in Figure 1): a visual filter (f_V) and an auditory filter (f_A). The filtered signals are then multiplied in two mirror symmetric subunits (S1, S2), after one of them is delayed by a second auditory-visual filter (f_{AV}). The outputs of the two subunits are then multiplied to produce the MCD_{CORR} output (Eq – A.4), whereas the difference of the two subunits produces the MCD_{LAG} output (Eq – A.5; and Figure 1). The MCD_{CORR} and MCD_{LAG} detector outputs are then mapped into behaviourally obtained psychometric data (SJ or TOJ task) via a probit link function:

$$F(Y_{SJ}) = Y'_{SJ} = \beta_{0,SJ} + \beta_{CORR,\,SJ} \,.\, MCD_{CORR} + \beta_{LAG,\,SJ} \,.\, MCD_{LAG} \qquad (1a)$$

$$F(Y_{TOJ}) = Y'_{TOJ} = \beta_{0,\,TOJ} + \beta_{CORR,\,TOJ} \,.\, MCD_{CORR} + \beta_{LAG,\,TOJ} \,.\, MCD_{LAG} \qquad (1b)$$

where Y denotes the psychometric function obtained in the SJ and TOJ tasks, respectively and F is the probit link function. The coefficients β_{CORR} and β_{LAG} are estimated from the psychometric

functions and are fitted for each data set (TOJ, SJ) separately, but assuming the same detector inputs.

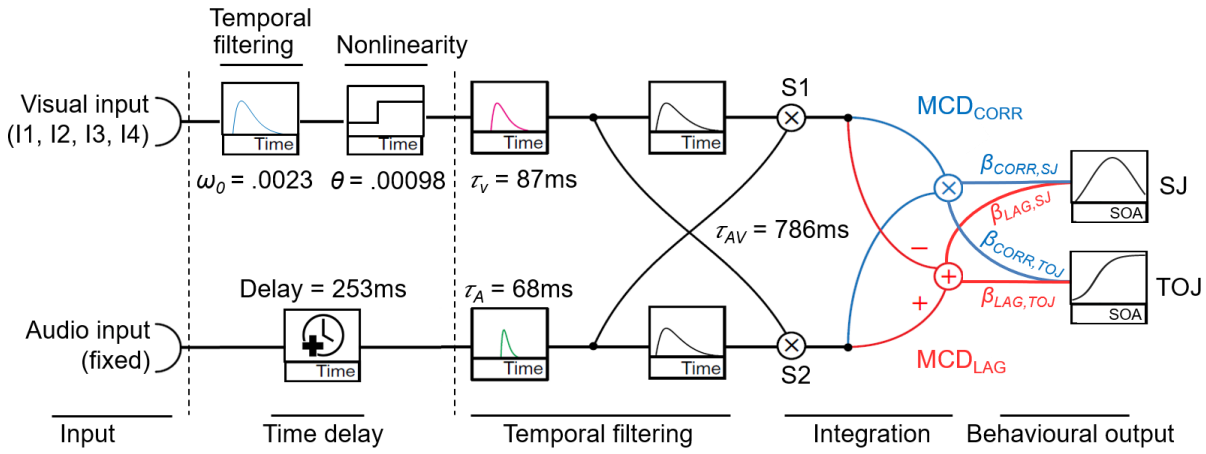

Figure 1. The Multisensory correlation detector (MCD) model (Parise & Ernst, 2016) with the added early visual-intensity dependent component (within dashed lines). The filter time constants (τ_V, τ_A, and τ_{AV}) were taken from the original MCD model and are not optimised for our data set. The parameters of the early component (filter cut-off ω_0 and threshold θ) were estimated using simple reaction time data from a separate experiment. The coefficients $\beta_{CORR,SJ}$ and $\beta_{CORR,TOJ}$ reflect the relative weightings of the MCD_{CORR} output for the SJ and TOJ data; similarly, $\beta_{LAG,SJ}$ and $\beta_{LAG,TOJ}$ denote the weights of the MCD_{LAG} detector (cf Eq 1). The four visual intensities are denoted by I1 (0.02cd/m^2), I2 (0.08cd/m^2), I3 (0.34cd/m^2) and I4 (1.34cd/m^2). Of the four models we compare, only models 3 and 4 include the early component (within the dashed lines). In models 2 and 4 the four intensities separately are fitted separately, with coefficients β optimised for each intensity; in models 1 and 3 the coefficients β are shared parameters (see Table1 and also Table C.1).

To model the known intensity-dependence of perceived simultaneity, we included an additional early processing stage (Temporal filter and subsequent nonlinearity; Figure 1), which converts variations in visual intensity into delayed stimulus onsets. This modified MCD model (MMCD) uses the filter time constants of the original MCD model (τ_V, τ_A, *and* τ_{AV}) and only the coefficients $\beta_{CORR,SJ}$ *and* $\beta_{CORR,TOJ}$ (Figure 1) are estimated from our behavioural TOJ and SJ judgements (see 3.1.2: Experiment). The parameters of the intensity-dependent early component (ω_0 *and* θ; Figure 1, left side) are estimated from a separate set of reaction time data (see section 2.3).

2.2. Variants of the MMCD model

To evaluate the modified MCD model, we fit four different nested variants of the model (Table 1). Model 1 does not allow for any intensity dependence of perceived simultaneity and the three coefficients β are estimated for all intensities simultaneously. In Model 2 the coefficients β are

Model no.	Early Processing Pars	Task	No. Pars (MMCD)	Visual Intensity	Estimated Parameters
1	-	SJ	3	I1,I2,I3,I4	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$ (shared parameters)
		TOJ	3	I1,I2,I3,I4	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$ (shared parameters)
2	-	SJ	12	I1	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I2	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I3	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I4	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
		TOJ	12	I1	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I2	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I3	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I4	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
3	ω_0, θ, n, t_{max}, RT_0	SJ	3	I1,I2,I3,I4	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$ (shared parameters)
		TOJ	3	I1,I2,I3,I4	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$ (shared parameters)
4	- ω_0, θ, n, t_{max}, RT_0	SJ	12	I1	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I2	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I3	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
				I4	$\beta_{0,SJ}$, $\beta_{CORR,SJ,}$, $\beta_{LAG,SJ}$
		TOJ	12	I1	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I2	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I3	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$
				I4	$\beta_{0,TOJ}$, $\beta_{CORR,\ TOJ,}$, $\beta_{LAG,\ TOJ}$

Table 1 –The 4 model variants. See text for details.

are optimised for each intensity level separately, resulting in 12 free parameters. for each task. Model 3 includes an intensity-dependent processing stage, which results in delay of the visual signal as a function of intensity. The coefficients β are shared parameters and fitted simultaneously for all

four intensity levels. Model 4 is like Model 3, but coefficients β are not shared across intensities. Model 3 is more parsimonious since fewer parameters are fitted (3 versus 12 free parameters for each task).

2.3. Characterising the early intensity-dependent component

We use previously measured simple reaction times to estimate the parameters of the early intensity-dependent processing stage, which converts a variation in intensity into a time delay. This delayed visual signal is then fed into the MCD correlation and lag detector, after further filtering (Figure 1). The reaction times data stem from a different experiment with 21 participants (9 F, 12 M), aged 19-49 (mean = 27.10, SD = 7.80, median = 25), using the same apparatus and identical stimuli to the current study. Simple reaction times were measured for unimodal stimuli of fixed intensities: at a single auditory intensity (15db above the mean estimated threshold) and at three visual intensities, 0.02 cd/m^2, 0.21 cd/m^2 and 1.34cd/m^2 (i.e. 6db, 15db and 24db above the mean estimated detection threshold). The median reaction times (RT) in ms for each individual observer as a function of intensity are shown in Figure 2 (black diamonds).

We used a simplified version of previous reaction time models (e.g. Cao, Zele, and Pokorny, 2007; Watson, 1986) since we are only interested in the relative intensity-dependent delay: an initial low pass filter, followed by an integrator drives a threshold crossing detector (see Appendix B1). The time from signal onset to reaching the response threshold directly predicts the stimulus dependent reaction time component (RT-$_S$) that is added to the irreducible RT (RT$_0$) to give the overall RT. We optimised five parameters: the filter order n of a Butterworth filter, the cut-off ω_0, the integration time t_{max} , threshold θ and the irreducible reaction time RT_0. (See table 2). The most relevant parameter is the number of filtering stages (n), as seen in Figure 2. A filter order of n=6 captures the linear relationship between reaction times and log intensity well. The cut-off ω_0 and the threshold θ cannot be estimated independently from each other since the absolute error is similar for different parameter combinations (see Figure B1). The time-limited integrator proposed by Cao et al. also has

little effect on the RT model fit provided the other parameters are optimised; t_{max} was set to the maximum duration of our signals, 100ms.

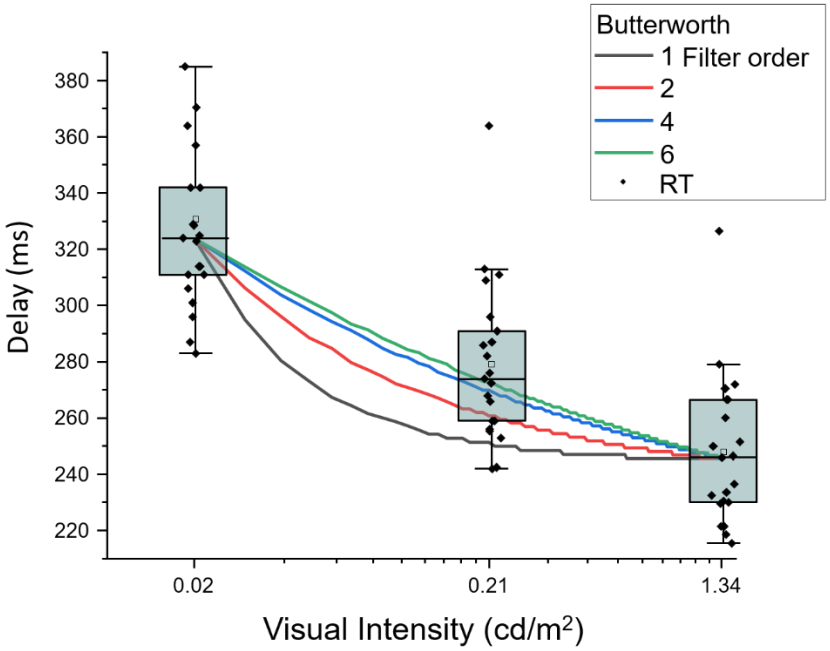

Figure 2: Median RT time data (black diamonds) are used to estimate the parameters of the Butterworth filter. Outputs for filters of order 1,2,4 and 6. The experimental data show individual RTs as well as median and upper/lower quartiles as box plots. The open symbols show the mean RTs. The optimisation used the median RTs (horizontal lines in boxes). The model output is quantised into milliseconds.

Butterworth filter	
Cut-off ω_0	0.0023 (f_c= 1.15Hz; f_s=1kHz)
order n	6
integrator t_{max}	100ms
threshold θ	0.00098
RT_0	207ms

Table 2: Filter and thresholding parameters. Note that there is no unique solution for the filter cut-off ω_0 and threshold θ since they co-vary (see also figure B1).

This time-delayed visual signal then feeds into the MCD (Figure 1). Because the visual signal is delayed by an intensity-dependent amount derived from the visual reaction times, the auditory

signal is also delayed by the reaction time measured for this particular auditory intensity (auditory RT=253msec). There is no free parameter for the alignment of the auditory and visual signals since it is determined by the respective reaction times.

2.4. The output of the Correlation and Lag detector as a function of intensity

The intensity-dependent delay of the visual signals (relative to the fixed delay of the auditory signal) is reflected in the relative location of the peak of the Correlation Detector and the zero point of the Lag Detector (Figure 3) for different visual intensities. The SOAs corresponding to the absolute peak and absolute minimum of the MCD_{CORR} and MCD_{LAG} detector outputs are also determined by the filter time constants of the visual and auditory filters (taken from the original MCD model and not optimised for our data set). In figure 3, each curve shows the respective output for a different intensity. The relative intensity-dependent shift of the MCD detectors (Lag and Correlation) determines the relative shift of the PSSs in the two tasks (see section 4.2)

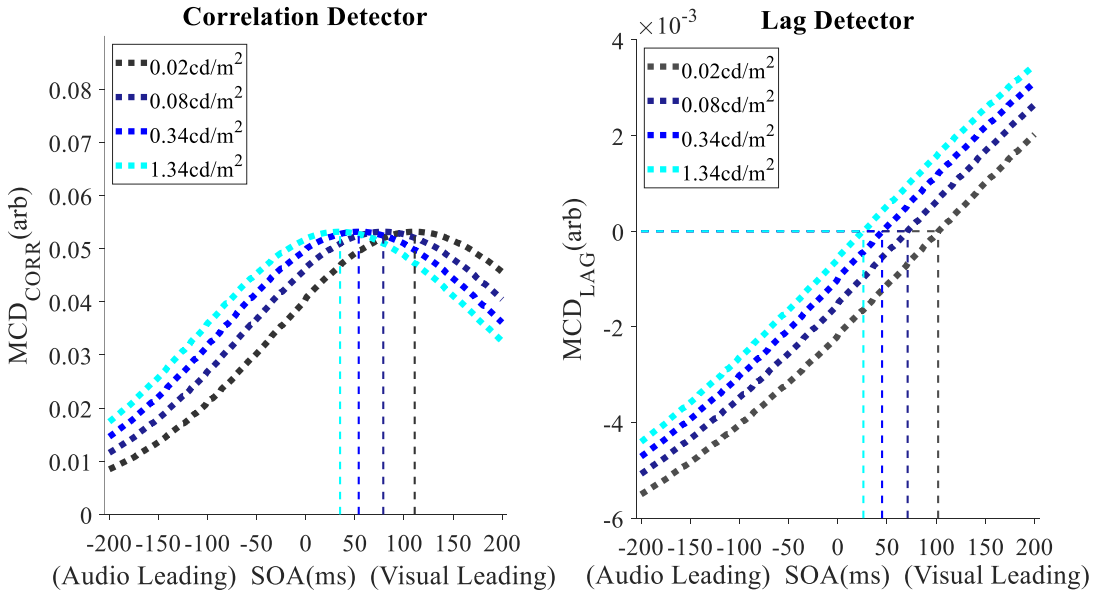

Figure 3. Outputs of the correlation (MCD_{CORR} - left) and lag detector (MCD_{LAG} - right) for the 4 visual intensities. The SOAs corresponding to the absolute peak and absolute minimum of the MCD_{CORR} and MCD_{LAG} detector outputs, respectively, are determined by the filter time constants of the visual and auditory filters (taken from the original MCD model and not optimised for our data set), along with the additional processing delays introduced by the early filtering and nonlinearity. The SOA corresponding to the peak of the correlation detector was 35ms (1.34cd/m2), 54ms (0.34cd/m2), 79ms (0.08cd/m2) and 111ms (0.02cd/m2). The SOA corresponding to a value of 0 in the lag detector was 26ms (1.34cd/m2), 45ms (0.34cd/m2), 71ms (0.08cd/m2) and 102ms (0.02cd/m2). The above

detector outputs were produced by models 3 and 4 which include the early intensity-dependent shift. For models 1 and 2, the MCD_{CORR} peaked at 43ms whereas the SOA corresponding to a MCD_{LAG} value of 0 was 35ms.

3. Experiments

The primary aim of our experiments was to test whether an early intensity-dependent processing delay predicts the intensity-dependence of perceived simultaneity in the both the SJ and the TOJ task. We therefore measured the point of subjective simultaneity (PSS) for four different intensities, using both the SJ and the TOJ task (the auditory level was fixed) and compared the empirical results with the predictions of the modified MCD model.

3.1. Methods

3.1.1. Preliminary experiments

In a preliminary experiment we estimated the visual and auditory thresholds of 8 participants (age range: 22-43, mean = 27.00, SD = 2.39, median = 24.50), 5 of which would later complete the SJ and TOJ tasks. Detection thresholds for auditory and visual thresholds were estimated using QUEST (Watson & Pelli, 1983) in a 2-IFC task. In the visual threshold estimation for each trial there were 2 clearly audible 100ms beeps, the onsets of which were separated by 1 second. A 100ms visual flash occurred 250ms after one of the two beeps (Koenig & Hofer, 2011). Participants were asked to press the left button if they believed the flash occurred after the first beep, and the right button if it occurred after the second. The auditory threshold estimation task mirrored that of the visual estimation, however participants were asked to detect a 100ms auditory beep which after one of two clearly detectable visual flashes. The mean auditory threshold was calculated as 20.54db and the mean visual threshold was 0.005 cd/m^2.

3.1.2. Simultaneity and Temporal Order Judgements

21 participants with normal hearing and normal or corrected-to-normal vision took part in the study. 3 of these participants were unable to perform the task reliably and produced a PSS in one or more conditions that was outside of our SOA (< or > 200ms) range and were therefore removed from subsequent analyses. Of the 18 participants, 10 were female (age range: 22-54, mean = 28.91, SD = 7.29, median = 27); 12 were naïve to both the SJ and TOJ tasks.

The experiment was conducted in a soundproof booth (IAC, Winchester, UK). Participants were seated 113cm away from the LED and loudspeaker which were at eye level for each participant. Stimuli were presented for 100ms and generated using a Tucker Davies RP2.1 real-time processor which controlled an LED and the loudspeaker. The LED was sealed in a plastic encasing behind 3 neutral density filters, with a fractional transmittance of 50%, 25% and 6.25%, respectively. Auditory stimuli were produced using a single 'Xenta M-219 Notebook speaker' which was located 1.62° below the LED. Participants were given a custom-built button box, with two large, shallow buttons to respond with. The responses and their timings relative to stimulus onset were recorded by the Tucker Davies RP system.

The visual flash was presented 6db, 12db, 18db and 24db above the mean estimated visual threshold, with a luminance of 0.02 cd/m^2 (scotopic light level), 0.08 cd/m^2, 0.34 cd/m^2 and 1.34cd/m^2 (mesopic light level). The auditory beep was presented 15db above the mean estimated auditory threshold at 1000hz and did not vary across conditions. The auditory signal was filtered using a raised-cosine filter to suppress transient clicks caused by stimulus onset and offset. The stimuli were presented at varying SOAs (-200ms, -150ms, -100ms, -50ms, 0ms, 50ms, 100ms, 150ms, 200ms), with negative values indicating that the auditory stimulus was leading. The inter-stimulus interval was 2000ms, plus a random value between 0-2000ms.

Stimuli were presented 36 times for each of the 9 SOAs and 4 intensity conditions, giving 1296 trials per participant, per task. These trials were presented in a random order across 6 blocks, with 6 trials at each combination of SOA and visual intensity within each block. Within each session, participants completed either the SJ or TOJ task; the four visual intensities that were presented in (pseudo)random order and the auditory sound level was held constant. The order in which the two tasks were completed was counterbalanced across observers.

Participants were instructed to hold the button box either vertically or horizontally, dependent on the task. For the SJ task, participants were asked to press the top button if they believed the two stimuli occurred simultaneously, and the bottom if they were non-simultaneous. For the TOJ task, participants were asked to press the left button if they detected a flash first, and the right button if they detected a beep first. They were then given two practice blocks, comprising of 36 trials of the highest visual intensity, split evenly across 3 SOAs (-200ms, 0ms and 200ms) in the SJ task, and 2 SOAs (-200ms and 200ms) in the TOJ task. The participant would only progress to the main experiment if they achieved at least a 66.7% correct response rate in the respective task in one of these two practice blocks. After feedback was given to the participant, they were given a 15-minute dark adaption period. The participant would then complete 6 blocks of the respective task, each separated by at least a 1-minute break. The two tasks were completed on separate days and participants were fully debriefed upon completion of both tasks.

3.2. Fitting the psychometric functions

To obtain the point of perceived simultaneity (PSS) in the SJ task, we fitted a gaussian function of the form:

$$a + b.\ exp(-0.5.((x - c) / d)^2) \quad (2)$$

The first parameter, a, is the offset along the y-axis, b is the amplitude of the Gaussian distribution, and c and d are the midpoint and width of the Gaussian distribution, respectively. To find the best-

fitting parameters, a nonlinear least-squares method was used (Function 'fit'; MatLab2018). The parameter c was used as the estimated PSS.

For the TOJ data, we fit a sigmoidal curve (Boltzman) with 4 parameters:

$$(a - b) / (1 + \exp((x - c) / d)) + b \quad (3)$$

The first two parameters, a and b, reflect the upper and lower asymptotes of the function. The 3rd parameter, c, is the location of the sigmoidal curve, defined as the midpoint between a and b, and d reflects the width of the curve. The SOA corresponding to a 50% 'light first' response rate was taken as the PSS.

4. Results

We first describe the intensity dependence of the PSS for both tasks and then the modified MCD model will be tested.

4.1. Psychometric functions for different intensity levels

The individual fits for each intensity level are shown in Figure 4 (a: SJ; b: TOJ). For the SJ task, a gaussian function was fitted (Eq 2 - dotted lines) to each observer's relative frequency (small dots in Figure 4a). The PSS was defined as the location of the curve (parameter c). The data of the TOJ task were summarised by fitting a sigmoidal Boltzman to the individual observer data (Eq 3 – Figure 4b).

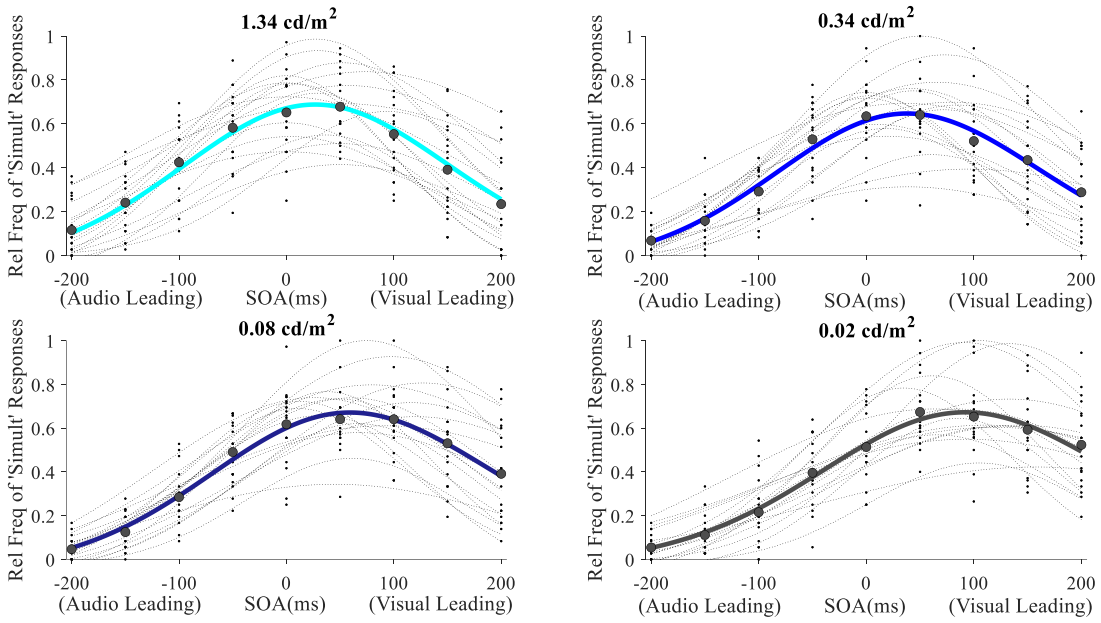

Figure 4a. Simultaneity judgement data. A gaussian function of the form f(x) = a + b . exp(-0.5 * ((*x - c*) / *d*)^2) was fitted (dotted lines) to each observers relative frequency (small dots). *Thick coloured lines represent the simulated curves using the mean parameters, and the larger circles represent the mean relative frequencies averaged across observers. The best fitting parameters (with SE) are as follows: 1.34cd/m^2: a = -0.067 (0.010), b = 0.77 (0.04), c = 31 (7), d = 133 (7). 0.34cd/m^2: a = -0.065 (0.009), b = 0.73 (0.04), c = 43 (7), d = 130 (7). 0.08cd/m^2: a = -0.062 (0.010), b = 0.76 (0.04), c = 65 (8), d = 136 (7). 0.02cd/m^2: a = -0.044 (0.010), b = 0.74 (0.04), c = 100 (8), d = 148 (9).*

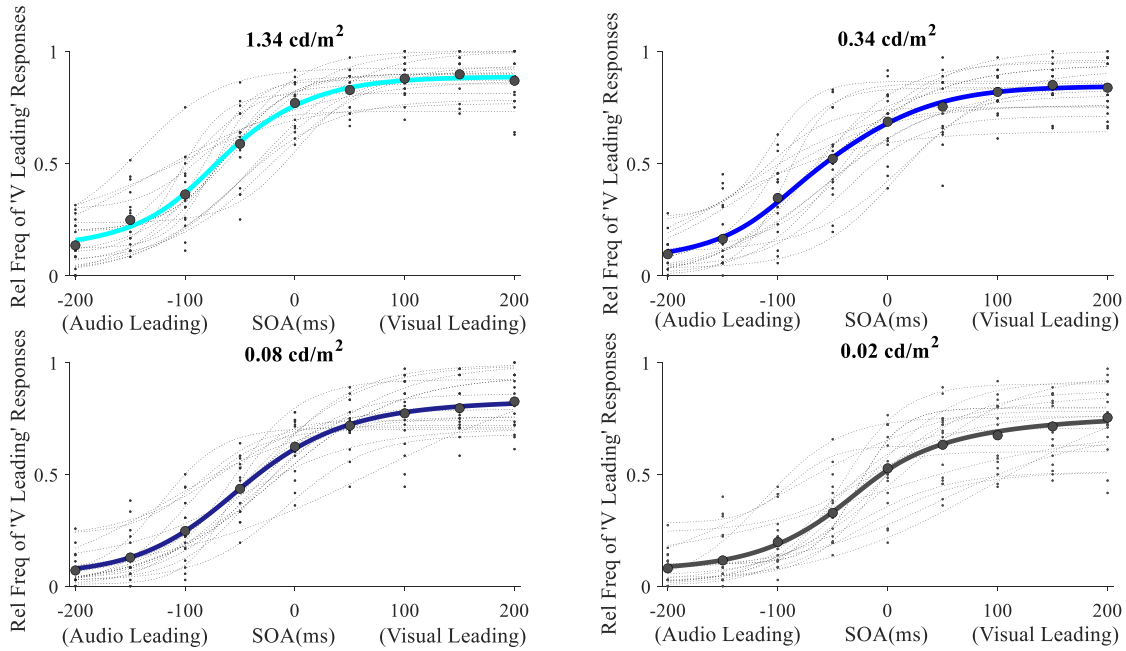

Figure 4b. Temporal order judgement data. A sigmoidal Boltzman function of the form f(x) = (a - b) / (1 + exp((x - c) / d)) + b was fitted (dotted lines) to each observers relative frequency (small dots). *Thick coloured lines represent the simulated curves using the mean parameters, and the larger circles represent the mean relative frequencies averaged across observers. The best fitting parameters (with SE) are as follows: 1.34cd/m^2: a = 0.116 (0.022), b = 0.89 (0.02), c = -72 (9), d = 38 (4). 0.34cd/m^2: a =*

0.054 (0.019), b = 0.84 (0.02), c = -71 (11), d = 42 (3). 0.08cd/m^2: a = 0.046 (0.017), b = 0.83 (0.03), c = -47 (11), d = 46 (4). 0.02cd/m^2: a = 0.063 (0.019), b = 0.76 (0.03), c = -26 (11), d = 45 (5).

Figure 5 shows the mean data for both tasks: while the absolute PSS shifts differ for the two tasks, the relative intensity-induced shift is very similar for the SJs and TOJs. For both tasks there was a significant main effect of intensity on PSS (SJ: $F(3,51) = 66.00$, $p < .001$, $\eta p^2 = .795$; TOJ: $F(1.48, 25.12) = 23.09$, $p < .001$, $\eta p^2 = .576$). Post hoc tests showed that all pairwise comparisons were significant at $p < .05$ or smaller p-values.

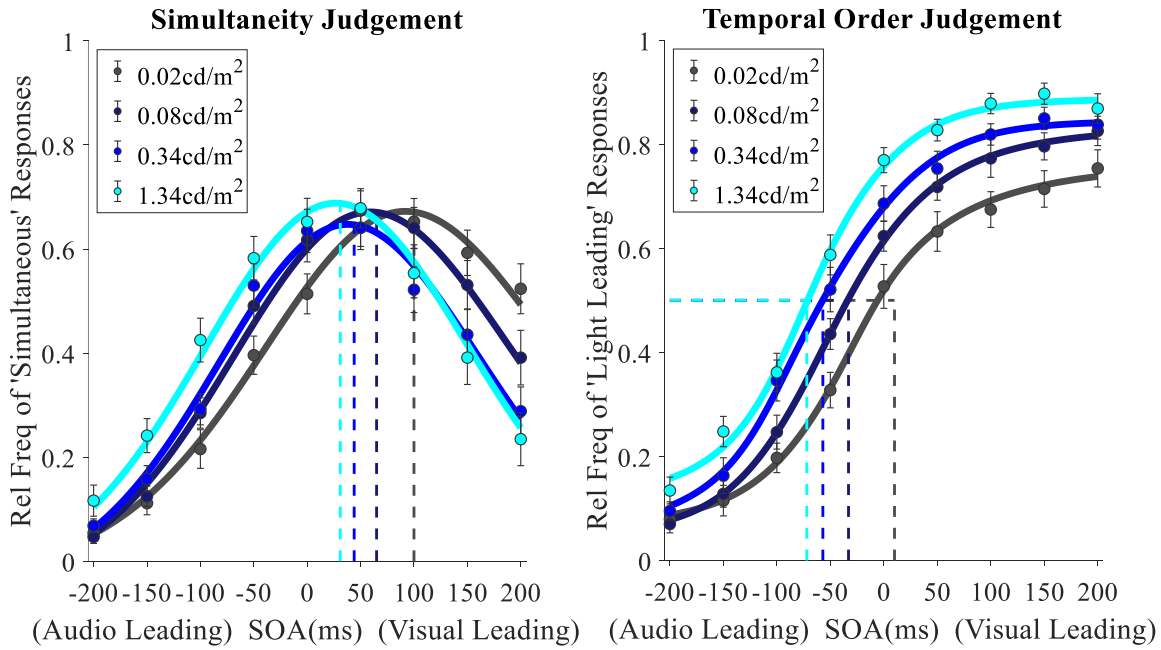

Figure 5. Mean fits (solid lines) for the SJ (left) and TOJ (right) task. Filled in circles represent mean relative frequencies and error bars represent the standard error of the mean. Dashed lines represent the mean PSS across all participants at each of the 4 intensities. The mean PSS in the SJ task was 31ms (1.34cd/m^2), 44ms (0.34cd/m^2), 65ms (0.08cd/m^2) and 100ms (0.02cd/m^2). The mean PSS in the TOJ task was -72ms (1.34cd/m^2), -57ms (0.34cd/m^2), -33ms (0.08cd/m^2) and 10ms (0.02cd/m^2).

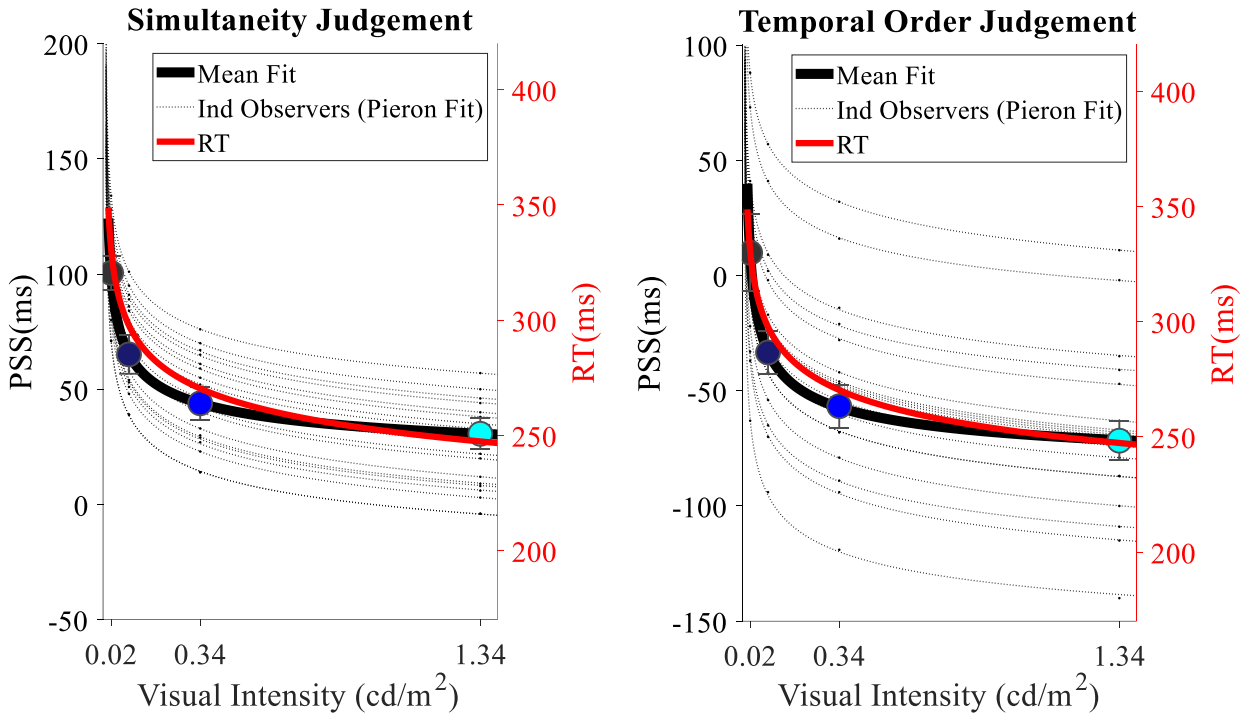

Figure 6. A Piéron function (RT (I) = a.(I^{-t}) + t_0) *fitted to the individual PSS estimates (left y-axis) for the SJ (left) and TOJ (right) task as a function of intensity. The Piéron fit for the simple RTs is overlaid (right y-axis). The mean best fitting parameters (with SE) are as follows for the SJ task: a* = 7.20 *(0.60); t = 0.58 (0.02), and* t_0 *= 30.87 (20.34).; for the TOJ task ; a* = 7.16 *(0.86); t = 0.58 (0.26), and* t_0 *= -60.76 (37.93). The mean of the individual PSS estimates are indicated by the large filled circles. Error bars represent the standard error of the mean.*

Comparing the psychometric functions and the PSS estimates as a function of intensity between the two tasks (Figure 6) shows that the dependence on intensity is very similar. It is notable, that the variation between observers is much larger in the TOJ task compared to the SJ judgement. Larger variability in the TOJ task has been observed previously (Recio, Cravo, de Camargo, & van Wassenhove, 2019) and may be related to the higher difficulty level reported for the TOJ task (Love et al., 2013).

4.2. Model evaluations

To compare the goodness of fits of the four models (Table 1, Section 2.2.) we used the corrected Akaike information criterion (AIC_c) which takes into account the number of free parameters (Anderson & Burnham, 2002). The corrected version of this statistic is recommended for sample

sizes <40 (Anderson & Burnham, 2002) and was calculated using the formula from Symonds and Moussalli (2011):

$$AIC_c = n[\ln(RSS/n)] + 2k + [2k(k+1)/(n-k-1)] \quad (5)$$

where n denotes the number of data points, k the number of parameters and RSS the residual sum of squares.

The AIC_c was calculated for each observer separately and averaged across observers (Bhardwaj, Van Den Berg, Ma, & Josic, 2016). The results of our model comparisons are summarised in Table C.1 (Appendix C) and Figure 8. Model 3 is the best model and over 4 times (4.76) and 5 times (5.34) more likely than the next best model (model 2) in the SJ and TOJ task, respectively. The predicted psychometric functions for model 3 are shown in Figure 7 together with the observed relative frequencies. The finding that Model 3 (3 parameter model) is the most likely model implies that the PSS shifts at different intensity levels can be explained by a common early processing delay which affects both tasks in a similar manner.

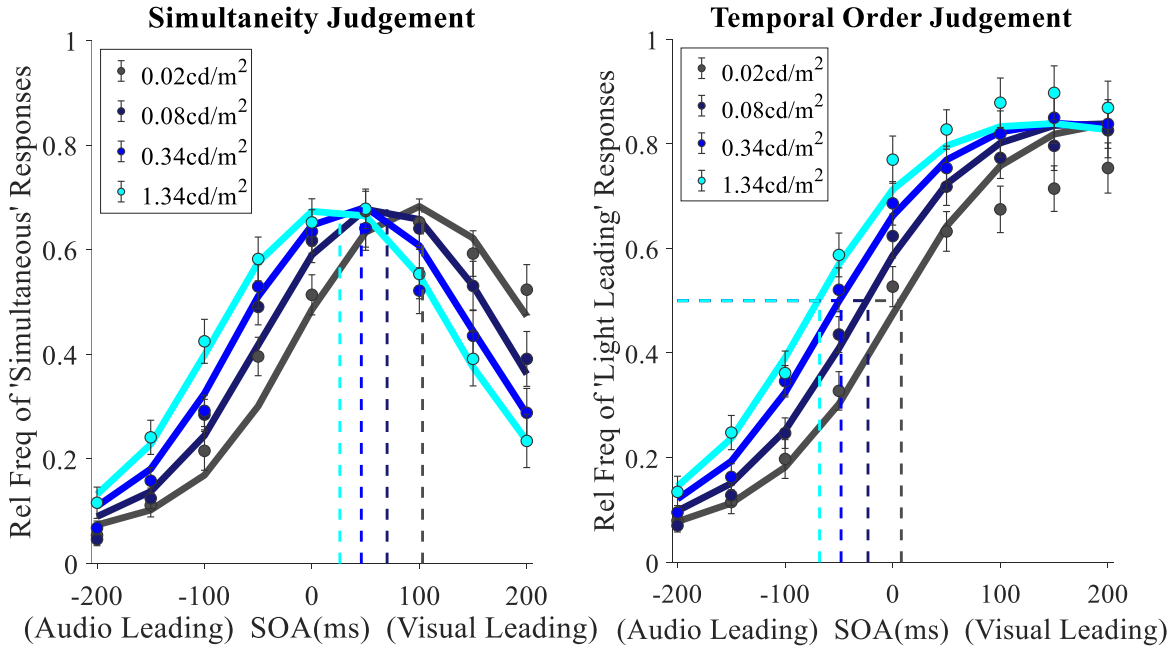

Figure 7: Behavioural data fitted using model 3. The solid lines represent the data fitted using the mean parameters, and the dashed lines represents the mean PSS when model variant 3 was fitted individually to each participant. The mean PSSs are 37ms (1.34cd/m^2), 45ms (0.34cd/m^2), 61ms

(0.08cd/m^2) and 101ms (0.02cd/m^2) in the SJ task and -55ms (1.34cd/m^2), -47ms (0.34cd/m^2), -31ms (0.08cd/m^2) and 8ms (0.02cd/m^2) in the TOJ task.

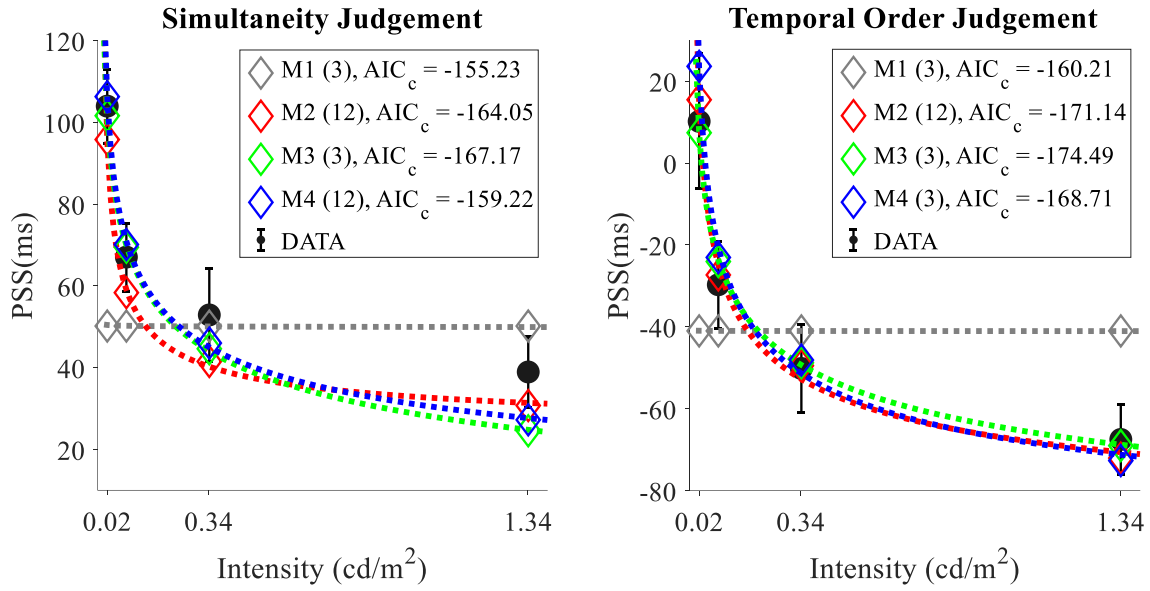

Figure 8. Comparison of the individual PSS estimates (+-SE for the SJ (left) and TOJ (right) tasks. Each colour refers to a different model prediction. Figure legend values in parentheses represent the number of fitted parameters for each model.

The predicted PSS values for all models (diamonds) are shown in Figure 8 together with the observed mean PSSs (discs with error bars). Model 1 (M1) does not allow for any intensity-dependency of the PSS hence predicts the same PSS for all intensities. Models 2 to Models 4 all predict the change in PSS as a function of intensity, with better accuracy for the TOJ task (right panel).

4.3. Correlation between the PSS obtained in the SJ and TOJ tasks

Finally, the PSS values obtained in the SJ and TOJ task (Figure 9 left) are not correlated (for 0.02 cd/m^2 ($r(16) = -.12$, $p = .648$), 0.08 cd/m^2 ($r(16) = -.14$, $p = .590$), 0.34 cd/m^2 ($r(16) = .02$, $p = .925$) and 1.34cd/m^2 ($r(16) = -.04$, $p = .861$). On the right side are the predicted PSS of Model 3 (Figure 9 right). No significant correlations between the predicted PSSs in the SJ and TOJ task were found at any intensity level (0.02 cd/m^2 : $r(16) = -.08$, $p = .749$; 0.08 cd/m^2 : $r(16) = -.08$, $p = .746$; 0.34 cd/m^2 : $r(16) = -.10$, $p = .691$; 1.34cd/m^2 : $r(16) = -.08$, $p = .740$). The degree of correlation depends on the relative

weights of the two internal detectors (cf Table D.1). Hence a lack of correlation is explained by observers assigning different weights to the Lag and Correlation detectors, when performing the SJ versus the TOJ task. This lack of correlation is found for all four intensity levels.

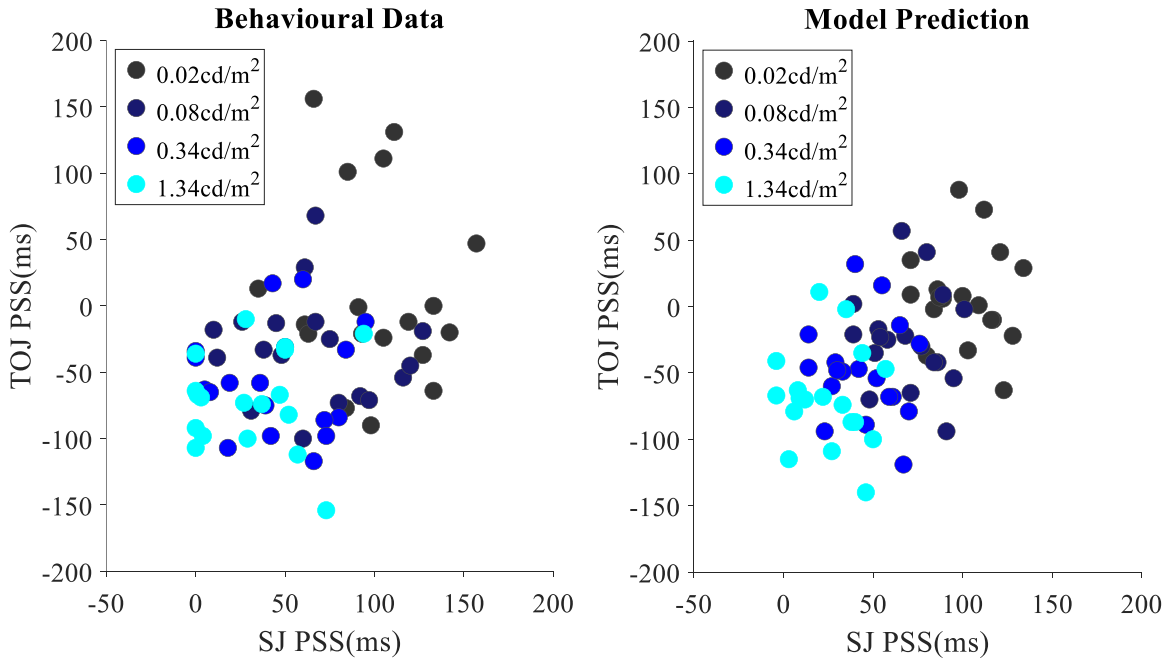

Figure 9. Scatter plot showing the lack of correlation between the SJ and TOJ PSS values.

5. Discussion

The main purpose of our study was to further characterise and model the intensity dependence of perceived simultaneity of auditory and visual events. Participants completed the SJ and TOJ task at 4 interleaved visual intensities, with a constant auditory intensity. The results showed a significant effect of intensity on both the SJ and TOJ task. Reaction time data from a separate experiment were used to estimate the parameters of a non-linearity added to the MCD model, and four variants of the modified MCD model (MMCD) were compared. The MMCD model provided a good fit to our data and is consistent with the idea that the shift in perceived simultaneity is predicted by early intensity-dependent processing latencies.

5.1. Perceived audio-visual simultaneity is affected by intensity

The PSS values for the SJ task are generally shifted towards visual-leading (Figure 5) in comparison to the TOJ task, where the PSS are shifted towards audio-leading, consistent with past findings (Van Eijk, Kohlrausch, Juola & van de Par, 2008). Crucially, the PSS across both the SJ and TOJ tasks showed a considerable intensity dependence: a reduction in visual intensity leads to the PSS being shifted towards visual-leading onsets, hence requiring an earlier visual signal to be perceived as simultaneous with the auditory stimulus of a fixed intensity. Intensity-induced shifts on the PSS have been demonstrated before in both SJ (Leone & McCourt, 2015) and TOJ tasks (Boenke, Deliano & Ohl, 2009; Leone & McCourt, 2015). Our data are consistent with the notion that low-level delays inherent in the processing of signals with different intensities account for shifts in perceived simultaneity. These processing delays, caused by intensity-dependent neural transmission latency shifts, are fundamental physiological constraints of this system that are not specific to this audio-visual integration mechanism. We argue that these fundamental constraints could account, at least in part, for the intensity dependence of perceived simultaneity. The relatively large temporal binding windows that are routinely described in the literature (Powers, Hillock, & Wallace, 2009; Stevenson, Zemtsov, & Wallace, 2012; Wallace & Stevenson, 2014) may also be a reflection of the inherent variability of the underlying temporal representations.

There are various factors that may contribute to the inter-task differences in absolute PSS values, again including task specific biases (Linares & Holcombe, 2014). A notable feature of our PSS estimates is that while the absolute PSS values differ between the two tasks, the intensity-dependence is very similar for the SJ and TOJ judgment (Figure 6). This is consistent with a common mechanism underlying the perceived simultaneity in both tasks and will be explored in the next section.

5.2. An early intensity-dependent processing delay predicts perceived simultaneity

We tested four different variants of the MMCD model (Figure 1), either with or without an early nonlinearity which delays the signal (see Table 1), and the weights of the Lag and Correlation detector (Figure 1) are either estimated separately for each intensity level (Model 2 and 4), or are shared parameters across all four intensities (Model 1 and 3). Model 1 served as a baseline since it did not allow for any intensity-dependence of the PSS. The most relevant models to compare are Model 2 (no assumption about intensity dependence is made and the beta coefficients are estimated for each of the four intensity level separately resulting in 12 free parameters for each task) with Model 3 (the relative shift in PSS is determined by the fixed early non-linearity and only 3 parameters are estimated for the four intensity levels for each task). The model predictions are very similar (see Figure 8 and Table C.1) but Model 3 is more the parsimonious model since only 3 parameters are estimated instead of 12. Figure 7 shows the predicted psychometric functions for model 3. Comparison with Figure 6 shows that the shift of the psychometric functions as a function of intensity is well captured by model 3 and supported by the model comparison using the corrected Akaike information criterion (Table C.1; Figure 8). Model 3 better captures the TOJ, in comparison to the SJ data, which is due to a similar dependence on visual intensity for RTs and the PSS in the TOJ task (see Figure D.1)

The early processing delay incorporated in the MMCD model (Model 3 and 4) was optimised based on reaction time measurements obtained with a different set of observers, but using the same apparatus and the same range of intensities. Consistent with previous studies (Piéron, 1913; Pins & Bonnet, 1996; Roufs, 1963), the relationship between unimodal visual reaction times and log intensity is linear in the intensity range tested in our experiments (Appendix D, Figure D.1, left panel). We observe a similar dependence on intensity in the SJ and TOJ tasks, as shown in Figure D.1 (middle and right panel): the observed shifts in perceived simultaneity are also approximately

linearly related to log intensity. The slope for the reaction times (-19.9) is slightly steeper than for the SJ task (-14.9), but very close to the slope in the TOJ task (-18.1).

Jaśkowski (1992) assessed the effect of visual intensity (from 0.4 to about 30 cd/m^2) on the PSS in the TOJ task and on simple reaction times and concluded that reaction times are reduced by a factor of 2 in comparison to perceived simultaneity, which is inconsistent with our finding that TOJ and reaction times show a similar intensity dependence. A major difference between the two experiments is that Jaśkowski (1992) varied the visual intensity across, rather than within experimental blocks, which may have introduced a task-specific response bias (Miyazaki, Yamamoto, Uchida & Kitazawa, 2006) or task-specific shifts in decision criteria (Yarrow et al., 2011; Linares & Holcombe, 2014).

While the original MCD model explicitly removes information on intensity and keeps only the onset of the signals, our analysis shows that the intensity dependency of perceived simultaneity can be accounted for by early intensity-dependent unisensory latency shifts, in line with Mansfield (1973), as opposed to an intensity-dependent integration mechanism. Our most parsimonious model (Model 3) predicts that these signals are utilised in the same way across intensities, i.e. no decisional factors interact with the employment of the two detector outputs, which is consistent with the 'one-system-two-decisions' approach (Cardoso-Leite, Gorea, & Mamassian, 2007).

5.3. Lack of correlation between the PSS obtained in the TOJ and SJ tasks

The lack of correlation between the PSS values obtained in the SJ and TOJ (van Eijk, Kohlrausch, Juola & van de Par, 2008) has been confirmed in our experiments and shown to hold across four intensity levels (Figure 8). Explanations for the lack of correlation between perceived simultaneity measured with different tasks (SJ and TOJ) propose differences in the decisional processes rather than divergent early processes (García-Pérez & Alcalá-Quintana, 2012; Love et al., 2013; Yarrow et al., 2011) . The modified MCD model is consistent with this approach by postulating that the lack of

correlation between the SJ and TOJ task is due to assigning different weights to early auditory-visual mechanisms, while allowing for common early processing strategies with a similar dependence on visual intensity.

5.2 Conclusions

Our behavioural data show an intensity dependent shift in perceived simultaneity that is in line with past research on simultaneity and temporal order judgements. The most parsimonious model absorbs the intensity dependence entirely into an early unimodal sensory processing stage, suggesting that the effect of intensity on PSS is dependent on early processing latencies that are common across both judgement tasks. Decisional elements do not factor into the intensity-driven PSS shifts, but instead underlie the lack of correlation between the PSSs , as per the original MCD model (Parise & Ernst, 2016). The updated model provides a *process model* accounting for both the of lack of correlation between different tasks as well as the intensity dependency of these audio-visual judgements.

Acknowledgements: Thanks to Cesare Parise for sharing his code with us.

Our behavioural data is available via Mendeley data:

https://data.mendeley.com/datasets/5742kchy9b/draft?a=44ef34eb-d15b-4b8b-bea6-ce0e70f60add

And will be made available on:

https://pcwww.liv.ac.uk/~sophiew/senses.htm

Declarations of interest: none.

RH was employed as a demonstrator (TA) by the Department of Psychology at the University of Liverpool.

Colour print preference: Online only.

Figure column preferences are as follows:

1 column – Figures 2 and A.1

2 column – Figures 1, 3, 4, 5, 6, 7, 8, 9 and D.1

References

Adrian, E. D., & Matthews, R. (1927). The action of light on the eye: Part I. The discharge of impulses in the optic nerve and its relation to the electric changes in the retina. *The Journal of physiology, 63*(4), 378–414. https://doi.org/10.1113/jphysiol.1927.sp002410

Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(6), 2244–2247. https://doi.org/10.1073/pnas.0407034102

Anderson, D. R., & Burnham, K. P. (2002). Avoiding Pitfalls When Using Information-Theoretic Methods. In *Source: The Journal of Wildlife Management* (Vol. 66).

Barnett-Cowan, M., & Harris, L. R. (2009). Perceived timing of vestibular stimulation relative to touch, light and sound. *Experimental Brain Research*, *198*(2–3), 221–231. https://doi.org/10.1007/s00221-009-1779-4

Bhardwaj, M., Van Den Berg, R., Ma, W. J., & Josic, K. (2016). Do people take stimulus correlations into account in visual search? *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0149402

Binder, M. (2015). Neural correlates of audiovisual temporal processing - Comparison of temporal order and simultaneity judgments. *Neuroscience*, *300*, 432–447. https://doi.org/10.1016/j.neuroscience.2015.05.011

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–282. https://doi.org/10.1007/s00221-004-2005-z

Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., & Morrone, M. C. (2009). Temporal mechanisms of multimodal binding. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1663), 1761–

1769. https://doi.org/10.1098/rspb.2008.1899

Cao, D., Zele, A. J., & Pokorny, J. (2007). Linking impulse response functions to reaction time: Rod and cone reaction time data and a computational model. *Vision Research*, *47*(8), 1060–1074. https://doi.org/10.1016/j.visres.2006.11.027

Cardoso-Leite, P., Gorea, A., & Mamassian, P. (2007). Temporal order judgment and simple reaction times: Evidence for a common processing system. *Journal of Vision*, *7*(6), 11. https://doi.org/10.1167/7.6.11

Carrillo-de-la-Peña, M., Rodríguez Holguín, S., Corral, M., & Cadaveira, F. (1999). The effects of stimulus intensity and age on visual-evoked potentials (VEPs) in normal children. *Psychophysiology*, *36*(6), 693–698. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10554583

Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological Review*, Vol. 122, pp. 232–241. https://doi.org/10.1037/a0038696

Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., & Molholm, S. (2011). Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. *Experimental Brain Research*, *213*(2–3), 167–174. https://doi.org/10.1007/s00221-011-2670-7

García-Pérez, M. A., & Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: A quantitative model. *Psychonomic Bulletin and Review*, *19*(5), 820–846. https://doi.org/10.3758/s13423-012-0278-y

Harrison, N. R., Wuerger, S. M., & Meyer, G. F. (2010). Reaction time facilitation for horizontally moving auditory-visual stimuli. *Journal of Vision*, *10*(14), 16–16.

https://doi.org/10.1167/10.14.16

Hassenstein, B., & Reichardt, W. (1958). *Functional structure of a mechanism of perception of optical movement*. 797–801.

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

Jaśkowski, P. (1992). Temporal-order judgment and reaction time for short and long stimuli. *Psychological Research*, *54*(3), 141–145. https://doi.org/10.1007/BF00922093

Jaskowski, P., Rybarczyk, K., Jaroszyk, F., & Lemanski, D. (1995). The effect of stimulus intensity on force output in simple reaction time task in humans. *Acta Neurobiologiae Experimentalis*, *55*(1), 57–64.

King, A. J. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus Studies of adaptive plasticity in unilateral hearing loss View project Inferring the dynamic networks behind neural responses View project. *Springer*, *60*(3), 492–500. https://doi.org/10.1007/BF00236934

Koenig, D., & Hofer, H. (2011). The absolute threshold of cone vision. *Journal of Vision*, *11*(1), 21–21. https://doi.org/10.1167/11.1.21

Leone, L. M., & Mccourt, M. E. (2015). Dissociation of perception and action in audiovisual multisensory integration. *European Journal of Neuroscience*, *42*(11), 2915–2922. https://doi.org/10.1111/ejn.13087

Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *I-Perception*, *4*(4), 213–228. https://doi.org/10.1068/i0532

Leone, L., & McCourt, M. E. (2012). The question of simultaneity in multisensory integration. *Human*

Vision and Electronic Imaging XVII, *8291*, 82910J. https://doi.org/10.1117/12.912183

Linares, D., & Holcombe, A. O. (2014). Differences in perceptual latency estimated from judgments of temporal order, simultaneity and duration are inconsistent. *I-Perception*, *5*(6), 559–571. https://doi.org/10.1068/i0675

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments. *PLoS ONE*, *8*(1), e54798. https://doi.org/10.1371/journal.pone.0054798

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*. https://doi.org/10.1523/jneurosci.07-10-03215.1987

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279. https://doi.org/10.1016/0010-0285(82)90010-X

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343. https://doi.org/10.3758/BF03203025

Miyazaki, M., Kadota, H., Matsuzaki, K. S., Takeuchi, S., Sekiguchi, H., Aoyama, T., & Kochiyama, T. (2016). Dissociating the neural correlates of tactile temporal order and simultaneity judgements. *Scientific Reports*, *6*. https://doi.org/10.1038/srep23323

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11543

Piéron, H. (1913). II. Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *L'année Psychologique*, *20*(1), 17–96. https://doi.org/10.3406/psy.1913.4294

Pins, D., & Bonnet, C. (1996). On the relation between stimulus intensity and processing time:

Piéron's law and choice reaction time. *Perception and Psychophysics*, *58*(3), 390–400. https://doi.org/10.3758/BF03206815

Powers, A. R., Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *Journal of Neuroscience*, *29*(39), 12265–12274. https://doi.org/10.1523/JNEUROSCI.3501-09.2009

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., … Belliveau, J. W. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *European Journal of Neuroscience*, *31*(10), 1772–1782. https://doi.org/10.1111/j.1460-9568.2010.07213.x

Recio, R. S., Cravo, A. M., de Camargo, R. Y., & van Wassenhove, V. (2019). Dissociating the sequential dependency of subjective temporal order from subjective simultaneity. *PLOS ONE*, *14*(10), e0223184. https://doi.org/10.1371/journal.pone.0223184

Roufs, J. A. J. (1963). Perception lag as a function luminance. *Vision Research*, *3*(8), 81–91.

Shore, D. I., Spry, E., & Spence, C. (2002). Confusing the mind by crossing the hands. *Cognitive Brain Research*, *14*(1), 153–163. https://doi.org/10.1016/S0926-6410(02)00070-8

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, *219*(1), 121–137. https://doi.org/10.1007/s00221-012-3072-1

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529. https://doi.org/10.1037/a0027339

Symonds, M. R. E., & Moussalli, A. (2011). *A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion*. https://doi.org/10.1007/s00265-010-1037-6

Ulrich, R, Rinkenauer, G., & Miller, J. (1998). Effects of stimulus duration and intensity on simple reaction time and response force. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 915–928. https://doi.org/10.1037//0096-1523.24.3.915

Ulrich, Rolf, Allan, L. G., Giray, M., Schmid, H., & Vorberg, D. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. In *Perception & Psychophysics* (Vol. 42).

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 and Psychophysics*, *70*(6), 955–968. https://doi.org/10.3758/PP.70.6.955

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: Temporal order versus simultaneity judgments. *Experimental Brain Research*, *185*(3), 521–529. https://doi.org/10.1007/s00221-007-1168-9

Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. *Neuropsychologia*, Vol. 64, pp. 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

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*(4), 1518–1531. https://doi.org/10.1016/j.concog.2011.07.003

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments.

Perception and Psychophysics, *67*(3), 531–544. https://doi.org/10.3758/BF03193329

Zampini, M., Shore, D. I., & Spence, C. (2003). Audiovisual temporal order judgments. *Experimental Brain Research*, *152*(2), 198–210. https://doi.org/10.1007/s00221-003-1536-z

Appendix A. The MCD Model

The original multisensory correlation detector model (Parise & Ernst, 2016) assumes that both visual and auditory stimuli pass through modality specific low-pass filters: a visual filter (f_V) and an auditory filter (f_A) with temporal constants of 87.30ms and 68.40ms, respectively:

$$F_{mod}(t) = t\ exp(-t/\tau_{mod}) \quad (A.1)$$

whereby τ represents as the modality-specific time constant of the filters (with which the signals are convolved), and t is time. The filtered signals are multiplied in two mirror symmetric subunits; (S1, S2), after one of them is convolved with a second multisensory filter (f_{AV}).with a time constant of 785.90ms

$$S1 = \{[S_A(t) * f_A(t)] * f_{AV}(t)\} . [S_V(t) * f_V(t)] \quad (A.2)$$

$$S2 = \{[S_V(t) * f_V(t)] * f_{AV}(t)\} . [S_A(t) * f_A(t)] \quad (A.3)$$

The signals of the two subunits are combined to produce two detector outputs. S_1 and S_2 are multiplied to produce MCD_{CORR} whereas the difference of the two subunits produces MCD_{LAG} .

$$MCD_{CORR} = S1 . S2 \quad (A.4)$$

$$MCD_{LAG} = -S1 + S2 \quad (A.5)$$

The MCD_{CORR} and MCD_{LAG} outputs are then mapped into behaviourally obtained psychometric data via a probit link function:

$$F(Y) = Y' = \beta_0 + \beta_{CORR} * MCD_{CORR} + \beta_{LAG} * MCD_{LAG} \quad (A.6)$$

where Y denotes the psychometric function obtained via the SJ and TOJ tasks and F is the probit link function. The coefficients β_{CORR} and β_{LAG} are estimated coefficients

Appendix B. Characterising the early intensity-dependent delay

Below is the Matlab code that computes the signal delay for a set signal amplitude; we assume a 100ms duration signal starting at time 0.

```
function del =  getdelay(amp);
% returns delay (ms) for step signal of given amp

% parameters from optimisation
omega0 = 0.0023; % fc cut-off frequency; fs sampling
%rate; omega0= fc/(fs/2)
thold = 0.00098; % threshold
intwindow = 100; % ms
filterorder = 6; % filter order
RTzero = 207;    % non-reducable part of RT

% get filter coefficients
[b,a]=butter(filterorder,omega0);

%create step function signal starting at sample 1
x=ones(1000,1)*amp; %signal
x(101:1000)=zeros(900,1); % set rest of signal to zero

%% filter, integrate, threshold
outdata = filter(b,a,x); % filter the signal
outdata=  movsum(outdata,intwindow); % integrator
del = find(outdata>thold,1)+RTzero; % find first
% threshold crossing and add RTzero
end
```

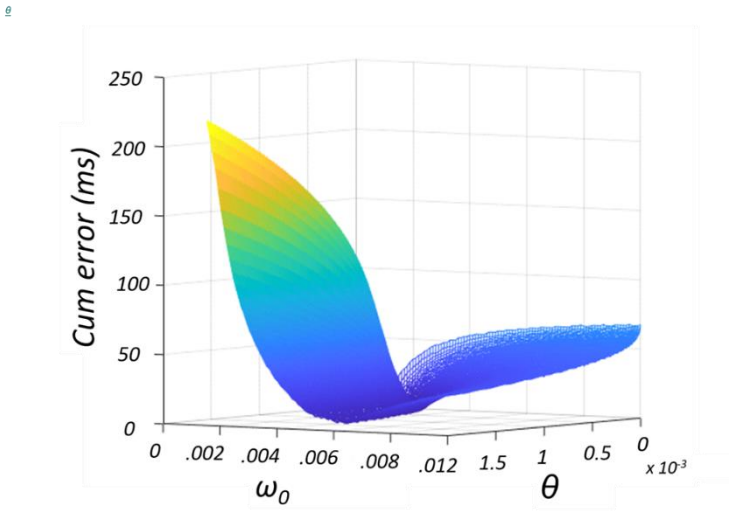

Figure B.1 – The effect of varying the cut-off and threshold on the cumulative error.

Appendix C.

Model no.	Pre-Processing?	Task	No. of Free Parameters	Visual Intensity (cd/m²)	β_0	β_{CORR}	β_{LAG}	PSS	RSS	AIC_c	ΔAIC_c	Akaike Weights	Evidence Ratio
1	No	SJ	3	0.02	-2.28	51.33	16.92	50.17	0.135	-155.23	11.94	0.002	0.003
				0.08					0.057				
				0.34					0.070				
				1.34					0.150				
		TOJ	3	0.02	-0.69	24.44	254.77	-41.00	0.159	-160.21	14.28	< 0.001	< 0.001
				0.08					0.051				
				0.34					0.064				
				1.34					0.116				
2	No	SJ	12	0.02	-1.91	42.76	160.00	95.78	0.037	-164.05	3.12	0.171	0.210
				0.08	-2.31	52.72	55.56	58.33	0.035				
				0.34	-2.58	56.90	-31.44	41.50	0.039				
				1.34	-2.83	63.81	-120.87	30.72	0.031				
		TOJ	12	0.02	-0.84	20.34	262.44	15.44	0.033	-171.14	3.35	0.151	0.187
				0.08	-0.72	23.73	287.30	-27.33	0.028				
				0.34	-0.57	24.75	276.78	-49.50	0.034				
				1.34	-0.11	20.60	315.90	-72.06	0.025				
3	Yes	SJ	3	0.02	-2.59	58.59	-79.87	101.61	0.090	-167.17	-	0.812	4.759
				0.08				69.44	0.065				
				0.34				44.61	0.064				
				1.34				24.56	0.083				
		TOJ	3	0.02	-0.50	23.72	242.36	7.39	0.101	-174.49	-	0.804	5.339
				0.08				-24.06	0.051				
				0.34				-48.78	0.053				
				1.34				-69.00	0.057				
4	Yes	SJ	12	0.02	-2.51	58.56	-83.82	106.28	0.049	-159.22	7.95	0.015	0.019
				0.08	-2.59	59.25	-77.88	70.11	0.044				
				0.34	-2.66	58.58	-72.78	46.00	0.042				
				1.34	-2.75	62.35	-83.22	27.22	0.029				
		TOJ	12	0.02	-1.00	30.57	138.97	23.61	0.038	-168.71	5.78	0.045	0.056
				0.08	-0.73	28.36	223.95	-23.06	0.032				
				0.34	-0.56	25.85	259.50	-48.11	0.035				
				1.34	-0.14	19.73	326.74	-72.67	0.025				

Table C.1 – Model comparisons. Values represent means across participants. ΔAICc, Akaike weights and evidence ratios represent comparisons within tasks. The relatively larger weighting (beta values) for the correlation detector (MCD_{CORR}) in the SJ task, and the lag detector (MCD_{LAG}) in the TOJ task are evident.

Appendix D.

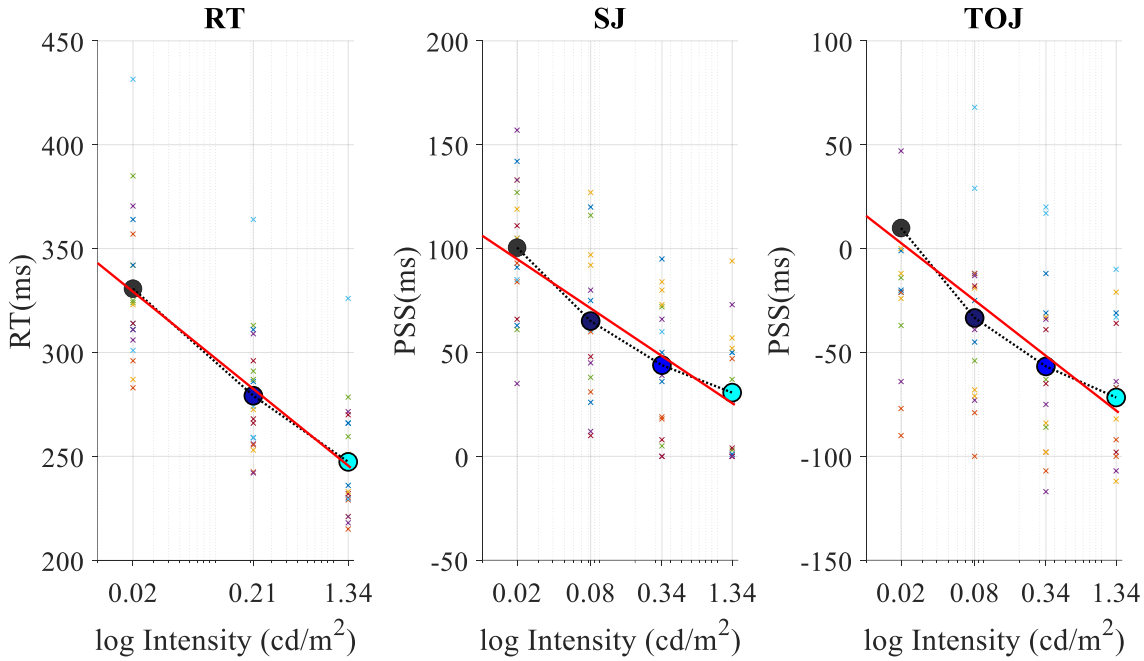

Figure D1- Reaction times and PSS values (for SJ and TOJ) are plotted as a function of log intensity. Individual observer data (x) and the mean data (o) are shown, together with the best linear-log fit (red line). The slope for the reaction times (-19.9) is steeper than for SJ (-14.9) and similar to TOJ (-18.1).