# Intensity-dependency and plasticity of perceived audio-visual simultaneity: behavioural and neural correlates

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by *Ryan Philip Horsfall*.

March 2021

# Table of Contents

# Abstract

The perception of simultaneity and temporal order of audio-visual stimuli is vital in understanding which sensory information emanates from the same physical events. Two tasks commonly used to assess the relative perceptual latencies of audio-visual stimuli are the simultaneity judgement and temporal order judgement tasks, both of which provide an estimate of the point of subjective simultaneity (PSS), which reflects the temporal stimulus offset (typically in milliseconds) where simultaneity is perceived. These two estimates, perhaps surprisingly, have been shown to be uncorrelated, leading to the suggestion that the two tasks are mediated by separate neural mechanisms. The present thesis addresses this possibility using electroencephalography (EEG), and demonstrates that task-based differences are evident only late in the event-related potential, attributable to differences in decisional, rather than early sensory, processes.

The PSSs estimated using these tasks are highly variable and depend on the relative intensity of the stimuli. By comparing various contender models, the author argues that the effect of intensity on the PSS can be explained by intensity-dependent processing latencies, and critically, that this intensity dependency is consistent across the two tasks. Furthermore, the current thesis highlights a novel process, whereby the PSS shifts depending on the visual intensity (as well as the leading modality) of the preceding trial, which is argued to reflect a decisional mechanism to compensate for differences in central arrival times caused by intensity-dependent processing latencies.

Finally, training observers using the SJ task has previously been cited as a potential method to improve the accuracy of audio-visual perceptual judgements in those with multisensory deficiencies. The presented data highlights the lack of generalisability of training induced

performance improvements between audio-visual stimuli with two visual intensities, highlighting limitations of these proposed interventions.

## Thesis outline

**Chapter 1 (introductory chapter)** briefly discusses the neural processes underlying the perception of audio, visual and audio-visual stimuli, and the influence of stimulus intensity on the processing of these signals.

**Chapter 2 (introductory chapter)** discusses the various tasks commonly used to estimate the influence of stimulus properties (such as temporal offsets and relative intensity) on perceived relative timing, with a focus on the simultaneity judgement (SJ) and temporal order judgement (TOJ) tasks. It discusses the various models of the processes underlying the SJ and TOJ tasks, the clinical importance of these tasks, and what they can tell us about the perceptual system.

**Chapter 3 (general methods)** describes the tailor-made equipment used throughout this thesis, then discusses and justifies the various methods implemented throughout the presented experiments.

Previous research has shown the PSSs estimated using the SJ and TOJ tasks to be uncorrelated. Furthermore, there is evidence of an inconsistent effect of visual intensity on the PSS and on facilitative multisensory integration, estimated using simple reaction times (RTs). **Chapter 4 (experimental chapter)** presents a pilot experiment, comprising three tasks (SJ, TOJ and RT), all of which involved identical audio-visual stimuli presented at two visual intensities. This aimed to replicate the effect of intensity on PSS, and to investigate the discrepancies reported between the tasks using a relatively large sample size.

The multisensory correlation detector (MCD) model accounts for the uncorrelated PSSs, by proposing that the tasks vary only at the decisional stage. However, this model cannot account for reported shifts in PSS that have been observed when relative stimulus intensity is manipulated. In **chapter 5 (experimental chapter)**, a modified MCD model with an early intensity-dependent processing stage is put forward, which is used to assess whether the previously reported PSS shift can be explained by intensity-dependent processing latencies.

The PSS has also been shown to dynamically shift, in response to the physical temporal offset of an immediately preceding trial. However, since physical temporal offsets of audio-visual stimuli do not typically fluctuate rapidly in the natural environment, it is unclear as to what the mechanism underlying this process exists to compensate for. **Chapter 6 (experimental chapter)** proposes that one of the main purposes of this process is to compensate for rapidly changing visual intensities of real-world stimuli. Through a combination of behavioural and electrophysiological methods, this chapter aimed to assess whether the perception of simultaneity is shifted, based on the intensity of an immediately preceding trial, and whether this hypothesised effect would be driven by the same underlying mechanism as that which underlies the PSS shift following physically (temporally) offset stimuli.

Perceptual learning can lead to a variety of improvements in perceptual capability and performance. Research has shown a large amount of variation in the stimulus properties (e.g., contrast or eccentricity) which perceptual learning induced improvements can be generalised across. Poor performance in the SJ task has been linked to various real-world implications, such as reduced speech perception, however this performance can be improved through training. **Chapter 7 (experimental chapter)** presents a between-subjects design, whereby observers completed the audio-visual SJ task, in response to trials with either bright or dim visual flashes, both before and after being trained to improve

performance at just one of these intensities. This experiment aimed to assess whether performance improvements at one visual intensity can be transferred to stimuli of a different intensity.

Finally, **chapter 8 (general discussion)** provides a general discussion of the presented experimental chapters which aimed to summarise how they contribute to the area in relation to past research.

**Notes on published materials**

It should be noted that chapter 5 and chapter 7 were formatted for their publication in the Journal of Mathematical Psychology and i-Perception, respectively. These remain identical to the published papers, though the page numbers, section numbers and figure numbers have been updated. Chapter 6 has been written and formatted for publication but has not yet been submitted.

The contribution of the PhD candidate (Ryan Horsfall) to the published material, consisted of the experimental designs, building of equipment, data collection, data analyses, production of figures and writing of the papers. The candidate's supervisors kindly contributed their knowledge and assistance in an advisory role throughout the contributions stated above.

**Consent from co-authors**

I give permission for the above papers to appear in the candidate's thesis.

Georg Meyer.

I also give permission for the above papers to appear in the candidate's thesis.

Sophie Wuerger.

# 1. Audio-visual perception

## 1.1. General introduction

Events in our immediate environment are rarely unimodal. The sight of a falling tree is accompanied by the sound of snapping twigs, just as the frightening sight of a fast-approaching car in a busy street will likely be paired with the screeching of tyres as it struggles to slow down.

The processing of this auditory and visual stimuli consists of multiple streams of sensory information, which converge to make a unified percept of the world around us. The integration of perceptual information also leads to a host of tangible benefits, such as improved detection rates of multisensory stimuli over their unimodal components (Barutchu et al., 2010) as well as reduced reaction times (Diederich & Colonius, 2004), potentially leading to the rapid detection and then avoidance of the aforementioned fast approaching car.

The accurate integration of auditory and visual information is, however, a highly complex process, and the incorrect binding of audio-visual information can lead to perceptual errors (e.g., Mcgurk & Macdonald, 1976). Such complexity is largely driven by the fact that the relative arrival times of auditory and visual information at 'multisensory' regions of the brain are highly variable. This is because auditory and visual modalities are detected by different sensory organs, and then processed in separate sensory pathways, eventually reaching primary sensory cortices located in separate lobes of the brain. Processing latencies vary between these two pathways, meaning that visual and auditory information often reach the primary sensory cortices at differing times. Furthermore, these latencies

vary based on the properties of the perceived stimuli, most critically for this thesis being their relative intensity.

This chapter provides a basic summary of the processing streams of auditory and visual information. It then briefly covers the various neural structures underlying the processing of audio-visual stimuli, and then discusses the various principles which dictate whether the information from the two modalities is integrated (at least, in the traditional sense). Finally, it discusses the effect of stimulus intensity on processing latencies.

### 1.2. Auditory and visual processing

For early visual processing, following the detection of light in the retina, retinal ganglion cells project signals to the lateral geniculate nucleus (LGN) in the thalamus, which then projects signals to the striate cortex (Michael, 1969). The striate cortex is organised as a retinotopic map, in that different parts of the visual field are mapped onto specific regions (and groups of neurons) (Hubel & Wiesel, 1959), analogous to the layers of the LGN (Le Gros Clark, 1940) and similar to the 'tonotopic' map in the primary auditory cortex (Merzenich, Kaas, & Roth, 1976; Reale & Imig, 1980). Information is projected from the striate cortex to the extrastriate cortex, a vast and complex network of regions spanning the parietal and temporal, as well as the occipital lobe (Felleman & Van Essen, 1991; Van Essen, 1990). Traditionally, it was believed that the visual information was processed entirely hierarchically in these cortices (e.g., Van Essen, 1990), however this has since been challenged, with the use of more modern neuroimaging techniques revealing a complex processing stream consisting of both bottom-up processes, and top-down processes which act on both 'early' and 'late' visual processing areas (Gilbert & Li, 2013).

In comparison, early auditory processing begins after sound vibrations which have reached the cochlea (e.g., Hallpike & Smith, 1934) are converted into action potentials (Pratt et al., 1992). This information travels along complex ascending pathways, arriving at the inferior

colliculus in the midbrain, and then the median geniculate nucleus in the thalamus, before reaching the primary auditory cortex in the temporal lobe (Antunes, Nelken, Covey, & Malmierca, 2010; Nelken et al., 2005; Rauschecker, 2015). Auditory information is then processed across a range of brain regions which form the 'auditory belt'. These include, (but are not limited to) parts of the superior temporal cortex, the prefrontal cortex and the inferior parietal cortex (Kaas & Hackett, 2000; Rauschecker, 2015). Again, this processing stream is not entirely linear, for example, there are descending pathways from the cerebral cortex to the auditory thalamus (for a review see He, 2003), and processing in the auditory cortex (Park, Ince, Schyns, Thut, & Gross, 2015) and even in the brainstem (Forte, Etard, & Reichenbach, 2017) are modulated by top-down processes (for a review see Lesicko & Llano, 2017).

Information from the auditory and visual modalities must be combined, in order to form an accurate percept of the surrounding environment. Importantly for this thesis, auditory and visual signals are initially processed through (partially) separate processing streams, and critically, these modalities are processed at different latencies. Auditory stimuli typically have shorter processing latencies than visual stimuli (Giard & Peronnet, 1999), and functional magnetic resonance imaging (fMRI) data highlights earlier activity in the primary auditory cortex, relative to the primary visual cortex (Alpert, Hein, Tsai, Naumer, & Knight, 2008; Raij et al., 2010) for suprathreshold stimuli. Furthermore, behavioural evidence shows that individuals typically react faster to auditory than visual stimuli (e.g., Pöppel, Schill, & von Steinbüchel, 1990). However, it should be noted that these processing and response latencies are intensity-dependent (see section 1.6).

### 1.3. Audio-visual processing

The brain is highly specialised for processing the multisensory information it receives, with a widespread 'multisensory' network involved in processing sensory information from multiple modalities (e.g., Ghazanfar & Schroeder, 2006). There are a range of neural

regions which show response modulation for a bimodal stimulus, in comparison to its unimodal components, including areas of higher-level processing (e.g., Barraclough, Xiao, Baker, Oram, & Perrett, 2005). For example, neurophysiology has revealed evidence of multisensory neurons (i.e. individual neurons which respond to stimuli of two or more modalities) in the cortex of cats (Wallace, Meredith, & Stein, 1992) and macaque monkeys (Mistlin & Perrett, 1990). Perhaps surprisingly, there is substantial evidence to show that neural activity in *primary* sensory areas is also modulated by multisensory stimuli (Besle, Fort, & Giard, 2004), including the auditory and visual cortices (Calvert et al., 1999). Using fMRI, Alpert, Hein, Tsai, Naumer and Knight (2008), investigated the neural pathway involved in the processing of simultaneous audio-visual stimuli. They reported initial activation in the primary visual and auditory cortices, which then propagates to the inferior frontal cortex at later latencies. Importantly, they also found that reductions in processing latency, observed for audio-visual, over unimodal stimuli, were widespread across both the inferior frontal cortex and the primary visual and auditory regions.

Consistent with fMRI data, EEG research has highlighted the processing pathway of an audio-visual stimulus, whereby the early activation seen in primary sensory areas later propagates to fronto-central electrodes (Giard & Peronnet, 1999). Importantly, early event-related potential latencies in occipital, occipital-parietal (<90ms) and temporal (<105ms) regions have all been shown to be modulated by multisensory stimuli, with an increase in P1 amplitude reportedly reflecting an increase in extrastriate activity in response to bimodal stimuli (Giard & Peronnet, 1999). Likewise, in the auditory cortex, there is evidence for multisensory interactions as early as 130ms in the left supratemporal auditory cortex (Klucharev, Möttönen, & Sams, 2003). Bi-lateral increases in amplitude of the auditory N1 (around 79-116ms) have also been reported (Giard & Peronnet, 1999), which are perhaps produced by the recruitment of additional neurons which are inactive for unisensory auditory processing (Giard & Peronnet, 1999). Finally, Molholm et al. (2002)

reported multisensory interactions as early as 46ms after stimulus onset, which they argued could be a result of direct connections between the primary visual and auditory cortices, or indirectly through feedback connections from a mediating multisensory region, such as the superior temporal sulcus. However, the methodology used to investigate this has since been criticised (McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2005), and remains controversial.

### 1.4. The principles of multisensory integration

The integration of multisensory information leads to a range of behavioural effects (see section 1.5), and is proposedly modulated by the convergence of multiple unimodal inputs onto individual neurons (Calvert et al., 1999). This theory has since been argued to be overly simplistic, since sub-threshold activation has been found in none 'multisensory' neurons (i.e. neurons which are activated by a unimodal visual, but not unimodal auditory stimulus), as a result of introducing an auditory stimulus alongside a visual stimulus (Allman et al., 2008). Still, this traditional view of multisensory integration has been widely studied, originating from the pioneering research of Meredith, Nemitz, and Stein (1987), who first detailed the importance of a single neural structure in this process that is located in the midbrain (Allen, 1927), namely, the superior colliculus (SC). Meredith, Nemitz and Stein's (1987) research using single cell physiology to investigate the effect of the properties of bimodal stimuli on action potential rates in the SC would then go on to form the principles of multisensory integration. These three principles are the spatial rule, the temporal rule and the principle of inverse effectiveness. As the temporal rule and the principle of inverse effectiveness are the most relevant to the current set of experiments, these are discussed the most thoroughly.

#### 1.4.1. The spatial rule of multisensory integration

The spatial rule states that the magnitude of multisensory integration (defined by the number of action potentials, their frequency and overall duration) is greatest when two

stimuli are spatially aligned. The integrative response of superior colliculus neurons is based on the overlap of unisensory receptive fields. The receptive fields of these neurons have a 'Mexican hat' disposition, leading to a large response when two stimuli come from an identical location, and a suppressed response when they are spatially offset within a specified region (Ursino, Cuppini, Magosso, Serino, & Pellegrino, 2009). The spatial rule is supported behaviourally, as spatially aligned stimuli have significantly lower detection thresholds (Meyer, Wuerger, Röhrbein, & Zetzsche, 2005) and faster reaction times (Gondan, Niederhaus, Rösler, & Röder, 2005) than spatially separate stimuli.

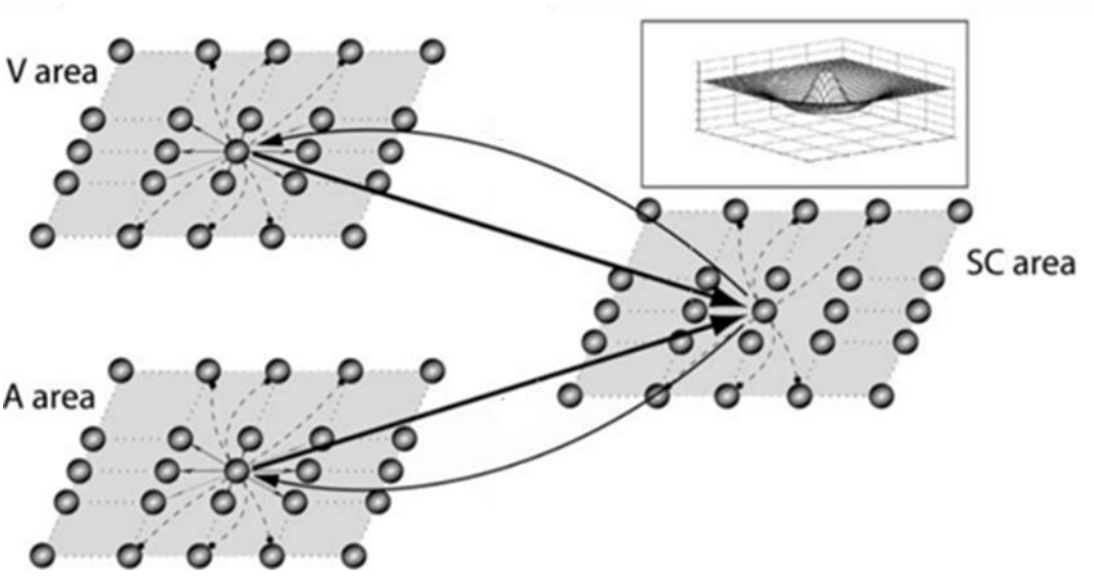


*Figure 1.1 - Schematic diagram displaying the effect of spatial alignment of visual (V area) and auditory (A area) inputs, onto multisensory cells in the SC (SC area). Each grey circle reflects an individual neuron, which are organised so that different parts of the external world are mapped onto specific regions of the 'grid'. Incoming auditory and visual signals cause an increase in action potentials in the respective unisensory neurons (dependent on the spatial location of the incoming stimulus), which then send excitatory signals to neurons in the SC, located in the same spatial position on the grid as each respective input. Each of these SC neurons has its own receptive field, which is organised in a 'Mexican hat' disposition (top right), so auditory and visual inputs from corresponding points on the grid*

*(i.e. in the V area and A area) will lead to the highest level of activity in the respective SC neurons. Figure adapted from Ursino, Cuppini, Magosso, Serino and Pellegrino (2009).*

### 1.4.2. The temporal rule of multisensory integration

The temporal rule states that multisensory integration will be greatest when two stimuli are temporally aligned (Meredith & Stein, 1986). While neurons in the SC show higher rates of action potentials in response to simultaneously presented bimodal stimuli than for unimodal stimuli (Rowland & Stein, 2007), a complete elimination of this response enhancement is evident for stimuli only hundreds of milliseconds apart (Meredith et al.,, 1987). Correspondingly, behavioural research has shown that race model violations (indicative of the integrated processing of the two modalities (Miller, 1982)) are only evident for temporally aligned (or very closely aligned) stimuli (Leone & McCourt, 2013). This finding is consistent with EEG data, where the latency of the N1 and P2 components have been shown to be dependent on the temporal relationship of the bimodal stimuli, with shorter latencies observed for synchronous, in comparison to asynchronous stimuli (Kopp & Dietrich, 2013). This reduction in auditory N1 and P2 latencies is evident when comparing bimodal stimuli to the sum of the unimodal auditory and visual ERPs, which is suggestive of activation that is specific for multisensory stimuli, as demonstrated using both speech (Van Wassenhove, Grant, & Poeppel, 2005) and non-speech stimuli (Kopp & Dietrich, 2013; Stekelenburg & Vroomen, 2007). Importantly, synchronous audio-visual speech is also associated with reduced auditory N1 and P2 amplitudes but introducing just a 200ms offset between the audio-visual stimuli eliminates this effect (Pilling, 2009). This effect is proposedly driven by increased activity in the superior temporal sulcus (Stevenson & James, 2009), through feedback connections to the auditory cortex (Ghazanfar, Chandrasekaran, & Logothetis, 2008).

### 1.4.3. The principle of inverse effectiveness

The principle of inverse effectiveness states that multisensory integration will be greatest when the two stimuli are of a lower intensity (Meredith & Stein, 1986). Since the seminal research by Meredith and Stein, the effect of stimulus intensity on multisensory integration has been widely studied, providing an array of findings which comply with this principle. For example, through isolating the 'multisensory' event-related potential (ERP) component of an audio-visual RT task, Senkowski, Saint-Amour, Höfle and Foxe (2011) showed that only low intensity stimuli produce an early latency (40-60ms) superadditive response in anterior and posterior regions, whereas medium and high intensity stimuli do not. This is also evident behaviourally, as reductions in simple reaction times (RTs) for an audio-visual bimodal stimulus, in comparison to the fastest of its unisensory components, is largest for low intensity stimuli (Diederich & Colonius, 2004).

Ursino, Cuppini, Magosso, Serino and Pellegrino (2008) used a neural network model of superior colliculus neurons to explain the process underlying the principle of inverse effectiveness. They demonstrated that whilst multisensory inputs always led to higher neural activity than their unimodal counterparts, this activity was superadditive for low-intensity stimuli, but subadditive for high-intensity stimuli. This was because for high-intensity stimuli, the activity of the superior colliculus neuron was near maximum, and therefore the addition of a second stimulus meant the increase in activity versus the unimodal stimuli was reduced.

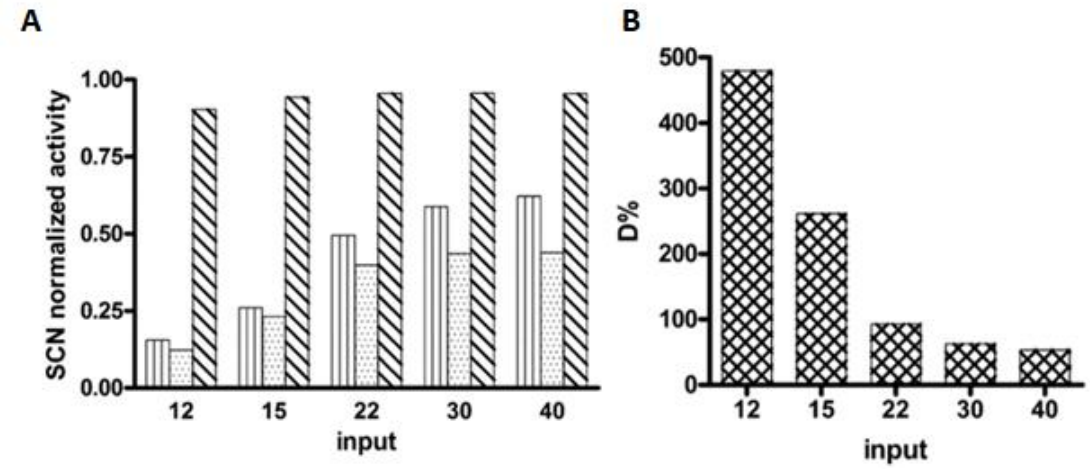


*Figure 1.2 - The effect of stimulus intensity on SC neuron activity. A) The activity of SC neurons, in response to visual (bar with vertical lines), auditory (bar with dots) and bimodal stimuli (bar with diagonal lines), across a range of intensities (values across x-axis which reflect normalised intensities based on the output of the SC in the model, and range between marginally above threshold (input 12) to saturation level (input 30 and 40)). While more intense stimuli lead to higher SC neuron activity (for both unimodal and bimodal stimuli), the percent increase of activity (see B) in response to a multisensory stimulus, in comparison to its unimodal components, is largest for low intensity stimuli. Figure adapted from* Ursino, Cuppini, Magosso, Serino and Pellegrino (2008).

### 1.5. Using the race model to assess multisensory integration

The greater firing rates in multisensory neurons (Meredith et al., 1987) as a result of the integration of signals from a bimodal stimulus is termed 'multisensory enhancement' (Meredith & Stein, 1986). The reduced reaction times to bimodal stimuli over their unimodal components (Diederich & Colonius, 2004; Leone & McCourt, 2013; Stevenson, Fister, Barnett, Nidiffer, & Wallace, 2012), termed the redundant signals effect (RSE), is widely reported in psychophysical research. However, the speeded response to multisensory stimuli is not necessarily indicative of a neural interaction when processing the two stimuli. A common method used to understand whether such an interaction

occurred is that of the race model inequality test (Miller, 1982). The race model states that in bimodal trials, the two stimuli are processed separately, and reaction times are dictated by the faster of the two channels. Race model violations occur when the RSE is too large to be attributable to the faster of the two unimodal stimuli; it can then be assumed that the two channels have been integrated at the neuronal level (Gondan & Heckel, 2008).

### 1.6. The influence of stimulus intensity on cell-to-cell communication

Action potentials occur when neurotransmitters bind to chemical-gated receptors of a postsynaptic neuron. This leads to the opening of ionic channels, increasing the membrane permeability of the cell to positively charged (e.g. sodium) ions (Rutecki, 1992). When the excitatory postsynaptic potentials from one or more presynaptic neurons depolarise the postsynaptic membrane past a certain threshold, voltage-gated channels will open, leading to the rapid depolarisation of a neuron's membrane potential, and consequently an action potential (for a review see Bean, 2007). Stimulus intensity has a known effect on the latency and rate of these nerve impulses (Podivinský, 1967). Increasing stimulus intensity reduces the latency of action potentials (Preswick, 1965), leading to a higher frequency of EPSPs at the postsynaptic neuron, thereby reducing the postsynaptic neuron's action potential latency (Curtis & Eccles, 1959).

The effects of stimulus intensity on processing latency can be demonstrated experimentally, for example, Gilliatt, Melville, Velate and Willison (1965) recorded the action potential latency between the median nerve at the wrist and upper arm. They stimulated this nerve at varying intensities and found reduced action potential latencies at higher voltages. More recent research has expectedly shown that these increases in intensity are also accompanied by the reduced latency of somatosensory evoked potentials (Shiga et al., 2001). The same rules apply to audio, visual and audio-visual perception, as increasing visual intensity reduces the latency of the visual evoked potential (VEP) (Carrillo-

de-la-Pena, Holguin, Corral, & Cadaveira, 1999; Lines, Rugg, & Milner, 1984) and increasing auditory intensity reduces the latency of the auditory evoked potential (AEP) (Jaskowski, Rybarczyk, & Jaroszyk, 1994). Similarly, increasing intensity leads to higher ERP amplitudes, evident in the increase in amplitude of the P1 component of the VEP (Brisson, Robitaille, & Jolicýur, 2007), reflecting increased activity of neurons in the extrastriate cortex (Clark, Fan, & Hillyard, 1994), and evident in the increase in the N1 and P2 components of the AEP (Adler & Adler, 1989).

Stimulus intensity has been shown to have a reliable effect on simple reaction times (RTs), a behavioural measure which, it has been argued, reflects perceptual processing latencies (Pins & Bonnet, 1997). Increased stimulus intensity leads to faster reaction times for both unimodal visual and auditory stimuli (Jaskowski, Rybarczyk, Jaroszyk, & Lemanski, 1995; Niemi, 1979; Ulrich, Rinkenauer, & Miller, 1998), as well as bimodal audio-visual stimuli (Harrison, Wuerger, & Meyer, 2010; Van der Stoep, Van der Stigchel, Nijboer, & Van der Smagt, 2016).

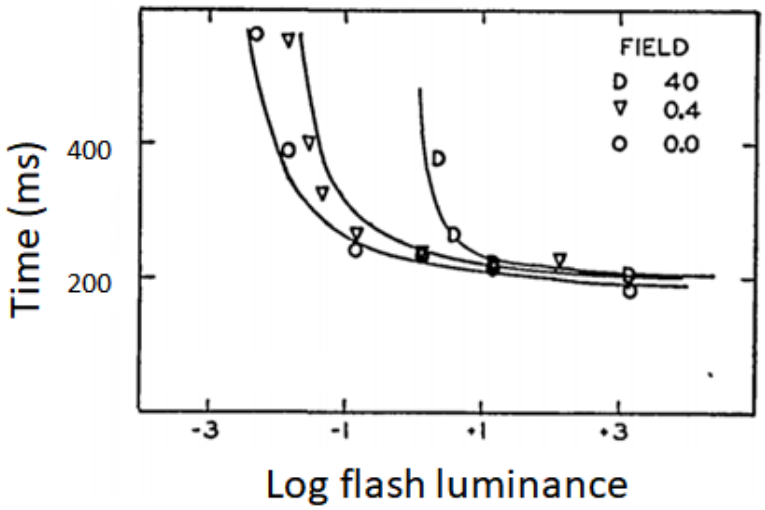


*Figure 1.3 - Median simple RTs of a single observer, in response to unimodal visual stimuli as a function of log flash luminance. Each curve represents one field (background) luminance. Figure adapted from* Bartlett and Macleod (1954).

Much like how simple RTs can be used to infer the effect of intensity on perceptual latencies, various perceptual judgement tasks allow us to discern the effect of intensity on *relative* perceptual latencies. For example, Roufs (1963) presented repetitive bimodal stimuli to just two observers, and asked them to adjust the offset of the two stimuli until they were perceived as simultaneous. By varying the intensity of the visual flash, while holding the auditory intensity constant, it was revealed the offset at which the visual flash had to be presented before the auditory beep, in order for simultaneity to be perceived, grew smaller as visual intensity increased. Since this early experiment, a wide range of similar tasks have been used to study relative timing perception, and its dependence on stimulus properties, including stimulus intensity. These will be explored in detail in chapter 2.

### 1.7. Chapter Summary

The current chapter provided a basic summary of the processes underlying audio-visual perception, alongside the principles which dictate whether stimuli from the two modalities are integrated. Furthermore, the chapter provided a basic summary of the effect of stimulus intensity on audio-visual processing latencies. This intensity-dependency is central to the current thesis, as each experimental chapter includes a manipulation of visual intensity, in an audio-visual stimulus pair. The following chapter discusses a wide range of literature on audio-visual processing, specifically in relation to audio-visual timing perception. Furthermore, it discusses the clinical importance of the tasks commonly used to assess audio-visual timing perception, and what these tasks can tell us about the perceptual system.

# 2.Chapter 2 –Audio-visual timing perception

## 2.1. Chapter Introduction

One of the most important indications of whether audio-visual stimuli were produced from the same source, is their temporal correlation (Nidiffer, Diederich, Ramachandran, & Wallace, 2018), and the importance of the co-occurrence of two stimuli in multisensory integration is well known (e.g., Stein & Meredith, 1993). However, due to the fluctuating properties of audio-visual signals experienced in the real-world which impact central arrival times, such as distance and stimulus intensity, audio-visual stimuli which occur physically simultaneously rarely reach multisensory regions of the brain at identical latencies. Judgements on the simultaneity and temporal order of audio-visual stimuli are therefore highly complex, but these judgements must be also accurate in order for perception to closely reflect the physical world. The processes underlying relative timing judgements have been widely studied; the relevant literature, and what it can tell us about the processes underlying audio-visual perception are summarised in this chapter.

## 2.2. Judgements of simultaneity and temporal order.

Many psychophysical paradigms have been used to understand the processes underlying audio-visual timing perception, including the sound induced flash illusion task (McGovern, Roudaia, Stapleton, McGinnity, & Newell, 2014), visual counting (Adams, 2016) and visual detection tasks (Li, Xi, Zhang, Liu, & Tang, 2019). The two tasks most relevant to the current thesis are the audio-visual simultaneity judgement (SJ) and temporal order judgement (TOJ) tasks. These tasks involve the presentation of visual and auditory stimuli across a

range of stimulus onset asynchronies (SOAs). While more complex versions exist (e.g., Van Eijk, Kohlrausch, Juola, & Van De Par, 2008; Yarrow, Martin, Di Costa, Solomon, & Arnold, 2016), the most basic two-alternative forced choice designs present participants with two options following each trial. In the audio-visual SJ task, participants are asked to press one button if they perceive the two stimuli to have occurred simultaneously, and another button if they perceive the two to have occurred separately. In the audio-visual TOJ task, participants are asked to press one button if they perceived the light to have occurred first, and the other button if they perceived the sound to have occurred first. It is possible to estimate a 'point of subjective simultaneity' (PSS) from the data of these two tasks (figure 2.1). For the SJ task, a function is fitted to the relative frequency of 'simultaneous' responses, and the SOA corresponding to the peak of this function is often taken as an estimate of the PSS (Barnett-Cowan & Harris, 2009; Leone & Mccourt, 2015). For the TOJ task, a function is fitted to the relative frequency of 'light first' responses, and the SOA corresponding to the 50% 'light first' response rate is typically used as the estimate of the PSS (Leone & McCourt, 2015). Through the manipulation of stimulus parameters, such as distance and eccentricity (Kopinska & Harris, 2004), intensity (Leone & McCourt, 2015), complexity (Noel, De Niear, Van der Burg, & Wallace, 2016; Van Eijk et al., 2008) and apparent causality (Kohlrausch, van Eijk, Juola, Brandt, & van de Par, 2013), it is possible to discern the influence of various factors on relative perceptual latencies.

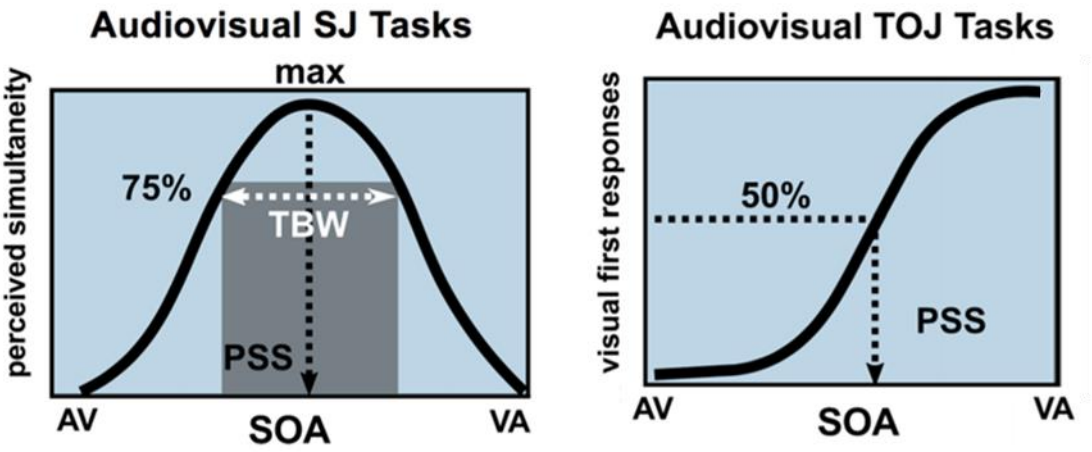


*Figure 2.1 – Typical mean fits (black, solid lines) for the simultaneity judgment (left) and temporal order judgement (right) tasks, and mean PSSs (black, dotted lines). The temporal binding window in this example (see section 2.3) is fitted to the 75% 'simultaneous' response rate (white dotted line). Figure adapted from Butera et al. (2018).*

Judgements on simultaneity and temporal order involve very complex processes, as transduction in the auditory system is typically quicker than the visual system (Pöppel et al., 1990; Recanzone, Guard, & Phan, 2000), but the relatively lower speed of sound (Recanzone, 2009) means two simultaneous stimuli will reach the respective peripheral sensors with a time delay that depends on distance to the source. Behaviourally, this is best highlighted with the evidence that there is a specific distance to an audio-visual source at which the reaction time to each modality (with specific stimulus properties) will be identical, which is referred to as the 'horizon of simultaneity' (Pöppel et al., 1990). Such an effect is also evident in the SJ task, whereby as the distance to an audio-visual stimulus increases, the time delay at which the auditory stimulus reaches the observer also increases (relative to the visual stimulus), thereby shifting the PSS towards the auditory-leading SOAs (Heron, Whitaker, McGraw, & Horoshenkov, 2007).

Despite the inherent complexity, humans have the ability to reliably combine multisensory stimuli emanating from a singular event. This is crucially important in the perception of

speech, as auditory speech signals are combined with accompanying visual lip movements to improve comprehension (Sumby & Pollack, 1954). The importance of this process is further highlighted by recent evidence that the perception of spoken sentences is most accurate when the audio and visual modalities are presented simultaneously, in comparison to unimodal visual or unimodal auditory conditions (Conrey & Pisoni, 2006). Furthermore, the incorrect fusion of visual lip movements and auditory phonemes leads to perceptual errors. For example, the 'McGurk effect' shows that when the visual lip movement of 'ga' is synched with the auditory sound of 'ba', the syllable 'da' is perceived (Mcgurk & Macdonald, 1976).

### 2.3. The temporal binding window

The simultaneity judgement (SJ) task allows for the estimation of the temporal window in which audio-visual stimuli are combined to form a singular percept. This is referred to as the 'temporal binding window' (TBW), and represents the range of SOAs at which two stimuli will be perceptually combined, resulting in a change of perception and behaviour (Noel, Wallace, Orchard-Mills, Alais, & Van Der Burg, 2015). The TBW is often calculated by taking the difference between the SOAs that correspond with the two points on the fitted curve that equal a 50% rate of 'simultaneous' responses (Stevenson, Zemtsov, & Wallace, 2012; Van Eijk et al., 2008), though other points on the curve are also sometimes used (e.g., Butera et al., 2018; Habets, Bruns, & Röder, 2017).

The TBW provides a temporal flexibility in what stimuli are integrated that is vital to ensure that physically simultaneous stimuli are perceptually bound. This is particularly important when the propagation speeds of these stimuli, both through the environment and the nervous system, vary based on the constantly changing properties of sensory signals experienced in the real-world (Murray & Wallace, 2011). Importantly, the size and shape of the TBW is far from static, and instead adapts to reflect the characteristics of the stimulus currently being perceived. For example, the size of the TBW varies based on stimulus

modality, as audio-visual stimuli are combined over a larger range of offsets than audio-tactile or visual-tactile stimuli (Noel et al., 2015). The difference observed between modality pairings proposedly exists because audio-visual stimuli occur at varying distances, meaning that variations in propagation time need to be considered. In contrast, the time between the production of a stimulus and it reaching the sensory system does not need to be considered for a tactile stimulus (Noel et al., 2015).

Interestingly, the shape of the audio-visual TBW has been shown to be asymmetric. The width of the visual-leading side of the TBW (typically estimated as the difference between physical simultaneity, and a specific point on the visual-leading side of the curve fitted to the data (see Cecere, Gross & Thut, 2016)) is typically reported to be larger than the audio-leading side, when measured using the SJ task (Cecere et al., 2016; Habets et al., 2017; Powers, Hillock, & Wallace, 2009; Stevenson, Wilson, Powers, & Wallace, 2013; Stevenson, Zemtsov, et al., 2012), though see Cecere, Gross, Willis and Thut (2017). This asymmetry is also evident for other perceptual tasks, as audio-visual offsets are more easily detected in an audio-visual film when the audio stimulus is leading (Dixon & Spitz, 1980), consistent with a reported asymmetry in the perception of the McGurk illusion (McGurk & MacDonald, 1976). This asymmetry evident in behavioural data is consistent with EEG data which highlighted the divergent neural networks underlying the audio-leading and visual leading TBWs (Cecere et al., 2017).

It is so far uncertain why the visual leading TBW is larger than the audio-leading side. It has been argued that this asymmetry may reflect the relative propagation times of stimuli through the environment, as a visual signal of a simultaneously occurring audio-visual event will always reach the observer before its auditory counterpart, which proposedly leads to an acquired tolerance of visual leading offsets (Powers et al., 2009; Stevenson et al., 2012). However, this explanation cannot account for the typically shorter processing

latencies in the auditory, over the visual system (King & Palmer, 1985; Pöppel et al., 1990; Raij et al., 2010). Perhaps a more compelling explanation derives from the alerting effect which an auditory stimulus has on the visual system about an upcoming visual signal (Thorne & Debener, 2014), which potentially reduces the size of the audio-leading TBW (Cecere et al., 2016).

There is an array of evidence to show that the TBW is flexible, based on both stimulus properties and the task at hand (e.g., Stevenson & Wallace, 2013). For instance, the size of the TBW increases with increasing stimulus complexity (Stevenson et al., 2014; Stevenson & Wallace, 2013) and with reduced stimulus intensities (Krueger Fister, Stevenson, Nidiffer, Barnett, & Wallace, 2016). The effect of intensity on TBW size is likely driven by the less reliable perception of temporal offsets for lower intensity stimuli (Krueger Fister et al., 2016) (i.e. there is an increased variability in the offset which is actually perceived for any given physical SOA), leading to an adaptation in TBW size to account for this uncertainty.

There is also evidence that the size of the TBW is plastic across the lifespan. However, unlike adaptations in size to best account for varying stimulus modalities and properties, the age-related plasticity is not necessarily beneficial. The audio-visual TBW reduces throughout childhood (Chen, Shore, Lewis, & Maurer, 2016; Lewkowicz & Flom, 2014) into adolescence (Hillock-Dunn & Wallace, 2012), and potentially continues to reduce in size up until just after 50 years old (Noel, De Niear, Van der Burg, & Wallace, 2016). Importantly however, its size increases again in older age (Bedard & Barnett-Cowan, 2016). This increase later in life is associated with perceptual errors, and increased fall risk (Setti, Burke, Kenny, & Newell, 2011). Thus, the size and shape of the TBW is clearly highly variable, based on both the properties of the perceived stimuli, and on individual differences between observers. Furthermore, this variability can contribute to real-world consequences.

Critically for the current thesis, while the TBW may increase in size in older age, this reduction in perceptual ability can potentially be mitigated by the experiences of the observer. For example, musical experience has been linked with reduced auditory N1 latencies (Behne, Sorati, & Alm, 2017) and reduced TBW size (Jicol, Proulx, Pollick, & Petrini, 2018; Lee & Noppeney, 2014). While it may be possible to utilise musical training to reduce the size of the TBW, more direct methods have been reported. These are discussed in section 2.12.1.

### 2.4. Clinical importance of SJ and TOJ tasks

The SJ and TOJ tasks consistently show considerable inter-individual variability (e.g., De Niear, Koo, & Wallace, 2016), that are dependent on observer-specific factors, such as gender (Zmigrod & Zmigrod, 2016), and abnormal performance on these tasks appears to be reflective of multisensory deficiencies caused by certain disorders. Importantly, it has been argued that these perceptual deficiencies could be used as an early marker for those in need of interventions (Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2011).

When measured using the SJ task, a larger than typical TBW is evident in those with autism (Foss-Feig et al., 2010; Stevenson et al., 2014) (though see Turi, Karaminis, Pellicano, & Burr, 2016), schizophrenia (Stevenson et al., 2017) and even obesity (Scarpina et al., 2016). This abnormality in multisensory processing has been shown to have notable impacts, as TBW size, estimated with both speech stimuli and simple flash-bleep stimuli, has been found to be negatively correlated with performance on a spoken sentence perception task (Conrey & Pisoni, 2006). Such a finding was replicated by Stevenson et al. (2018), who showed that the relationship between TBW size and speech perception performance, in this case measured using the speech in noise task, was mediated by the individual's proneness to the McGurk illusion. In other words, if an individual incorrectly binds discrete audio-visual information, then this reduces their ability to correctly separate visual lip

movements and auditory speech sounds, therefore negatively impacting their perception of audio-visual speech (Stevenson et al., 2018).

The steepness of the TOJ slope provides a measure of sensitivity to audio-visual offsets, that is often compared to the TBW estimated in the SJ task (Van Eijk et al., 2008). The TOJ task has clinical significance, as those with autism spectrum disorder (ASD) showed reduced sensitivity (i.e. reduced steepness of the slope fitted to the TOJ data) to audio-visual offsets in the TOJ task (de Boer-Schellekens, Eussen, & Vroomen, 2012), as did individuals with dyslexia (Hairston, Burdette, Flowers, Wood, & Wallace, 2005), obesity (Scarpina et al., 2016) and past concussion (Wise & Barnett-Cowan, 2018). Furthermore, the difference between those with ASD and typically developing observers is consistent across both flash-beep and audio-visual speech stimuli (de Boer-Schellekens et al., 2012). It should be noted however, that a recently published experiment reported no difference in TOJ performance between a typically developing and an ASD group (Kawakami et al., 2020), though the authors suggested that this could be due to the relatively small sample size.

While the SJ and TOJ tasks appear to be very similar, the estimates from these tasks are often highly contrasting (see section 2.5). Despite this, the SJ and TOJ tasks have frequently been implemented “interchangeably” (Keetels & Vroomen, 2011, para. 6), or in combination, to study varying disorders such as schizophrenia (Capa et al., 2014) autism (Regener, Love, Petrini, & Pollick, 2014) and obesity (Scarpina et al., 2016). It is therefore vital that the differences and similarities between the tasks are understood. The fact that data from the two tasks were combined in a recent meta-analysis on audio-visual integration and autism spectrum disorders (Feldman et al., 2018), emphasises the importance of exploring the degree of overlap in their underlying processes, as well as the estimates they provide. These differences will be discussed in detail in the following section.

### 2.5. Discrepancies in estimates from the SJ and TOJ task

Despite the methodology of the SJ and TOJ tasks varying in instructions alone, recent studies have highlighted various discrepancies in the estimates they provide. A common finding is that there is no correlation in the PSSs estimated using the audio-visual SJ and TOJ tasks (Basharat, Adams, Staines, & Barnett-Cowan, 2018; Binder, 2015; Linares & Holcombe, 2014; Love, Petrini, Cheng, & Pollick, 2013; Van Eijk et al., 2008; Vatakis, Navarra, Soto-Faraco, & Spence, 2008), a disparity that is evident when estimated using both simple and complex stimuli (Van Eijk et al., 2008). This discrepancy between the tasks is further highlighted by research which showed that behavioural improvements induced from training observers at one task, led to no noticeable improvements in the other task, despite the use of identical stimuli (Matthews, Welch, Achtman, Fenton, & FitzGerald, 2016). The reason why the two tasks would produce uncorrelated findings is so far uncertain; recent arguments have attributed this to variations at the decisional stage of processing (Parise & Ernst, 2016), whereas others have argued that the uncorrelated findings are indicative of separate processing mechanisms underlying the tasks (e.g., Zampini, Shore, & Spence, 2003; Love et al., 2013). These arguments will be explored in more detail in section 2.7.

Contrasting with the idea that the tasks may draw upon separate neural mechanisms, it has been proposed that task differences in PSS estimates can be explained by post-perceptual response biases (Machulla, Luca, & Ernst, 2016). Linares and Holcombe (2014) suggested that the uncorrelated PSSs reported in their data were a result of different biases associated with each task. These include a tendency to favour a certain response when participants felt uncertain, causing a task-specific shift in the PSS in the TOJ task. In comparison, such a bias in the SJ task would have no effect on the estimated PSS, and would instead change the amplitude of the fitted curve.

Alternatively, the divergent findings may also be explained by differences in decisional processes (e.g., García-Pérez & Alcalá-Quintana, 2012, 2015). Yarrow, Jahn, Durant and Arnold (2011) contended that judgements of simultaneity and temporal order are both made by comparing the relative offset of audio-visual signals to task-specific decisional criteria that are placed on an internal 'timeline' (see section 3.7 for more detail). For the SJ task, simultaneity is perceived if the relative latency differential falls between two criteria, whereas in the TOJ task, temporal order is determined based on the same latency differential in comparison to a single criterion. The differences in the placement of criteria for these two tasks provides a reasonably parsimonious explanation for their divergent estimates.

### 2.6. Neural basis of Audio-visual timing perception

The evidence that the PSS estimates obtained from the SJ and TOJ tasks are uncorrelated opens up interesting questions on the neurocognitive mechanisms underlying the tasks. Is it possible that the slight variation in instructions that separate the tasks leads to the brain drawing upon entirely separate mechanisms when processing and/or making judgements on *identical* stimuli? While task variations have been shown to modulate neural processing (e.g., Besle et al., 2004), it would perhaps be surprising if the SJ and TOJ tasks, which on the surface appear so similar, differ at such a fundamental level.

Neuroimaging techniques can provide unique insights into the mechanisms underlying the SJ and TOJ tasks to potentially elucidate whether they are divergent (e.g., Love et al., 2013) or shared (e.g., Parise & Ernst, 2016). The high spatial resolution of functional magnetic resonance imaging (fMRI) in particular, has the capability to assess whether the two tasks draw upon separate neural structures. Using fMRI, Binder (2015) reported shared activation across the audio-visual SJ and TOJ tasks in the parietal lobe. This region has previously been highlighted as having a crucial role in relative timing perception tasks, as those with parietal malfunctions have notable performance deficits (Husain, Shapiro,

Martin, & Kennard, 1997; Rorden, Mattingley, Karnath, & Driver, 1997). Binder's (2015) experiment also highlighted shared activity in, amongst others, the inferior frontal gyrus (IFG), which has been implicated in multisensory perception (Li, Seger, Chen, & Mo, 2020; Renier et al., 2009), as well as in the occipital lobe, which is vital in visual perception (e.g., Wandell, Dumoulin, & Brewer, 2007). The TOJ condition was associated with higher blood-oxygen-level-dependent (BOLD) signal in several areas of the left hemisphere, including the IFG and parietal regions. Finally, there was TOJ specific activation in the occipito-temporal junction, suggesting some differences in the neural regions required for each task. Similar research using tactile stimuli has supported the existence of independent neural activity that was specific for each task, though again, this highlighted the extra processing required to make temporal order judgements (Miyazaki et al., 2016).

In contrast to Binder (2015) and Miyazaki et al. (2016), a comparable experiment found that while the TOJ task showed higher transient BOLD activation, the neural structures underlying the two audio-visual tasks were entirely overlapping (Love, Petrini, Pernet, Latinus, & Pollick, 2018). Since BOLD signal is highly dependent on task difficulty in perceptual tasks (Culham, Cavanagh, & Kanwisher, 2001; Gould, Brown, Owen, Ffytche, & Howard, 2003), the influence of the subjectively higher difficulty of the TOJ, over the SJ task (Love et al., 2013) on these findings is unclear. Taken together with the high chance of type-1 error in fMRI analyses, as a result of the high rate of statistical comparisons (Bennett, Miller, & Wolford, 2009), the neuroimaging data in favour of separate mechanisms underlying the tasks should be interpreted with some caution.

The increased temporal acuity of EEG makes this method ideal for investigating the time course of the neurocognitive processes underlying the tasks, and to elucidate whether any differences relate to early, low level sensory processes or late decisional factors (though this has so far been underutilised). Basharat, Adams, Staines and Barnett-Cowan (2018)

were, to the best of the author's knowledge, previously the only individuals to compare the SJ and TOJ tasks using EEG. They reported a significant effect of task on the amplitude of the visual P1 component (recorded at a single occipital electrode), which they argued reflected different neural mechanisms underlying the two tasks. Furthermore, they reported a sustained N1 response over larger offsets for ageing adults versus younger adults, which they argued reflected the larger TBW estimated for older observers (Bedard & Barnett-Cowan, 2016; Noel et al., 2016). However, it must be noted that their analysis focused only on early ERP components, meaning that it is currently unclear whether the sole differences are in the early ERP, or whether there are also later differences between the tasks.

### 2.7. Models of simultaneity and temporal order judgements

There are several models which have been used to explain the mechanisms underlying SJs and TOJs. Ulrich, Allan, Giray, Schmid and Vorberg (1987) proposed a simple threshold model, where a central mechanism compares the arrival times of two stimuli, and if the latency difference between the signals is large enough then temporal order can be discerned, and if not, simultaneity is perceived. This model however, with its single common mechanism, cannot explain the lack of correlation commonly reported between the PSS estimates of the two tasks. Furthermore, it cannot account for the evidence that an observer does not necessarily perceive two stimuli as simultaneous when the temporal order cannot be discerned (Weiß & Scharlau, 2011).

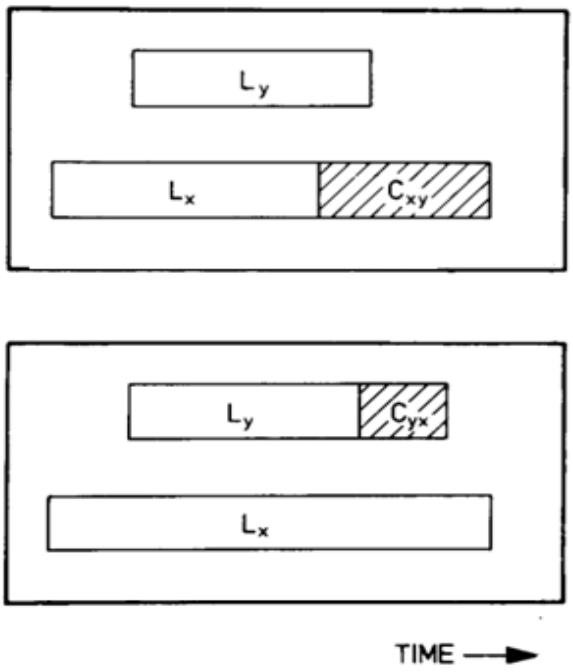


*Figure 2.2 – Schematic diagram of the threshold model. $L_x$ and $L_y$ represent the latency of an auditory and visual signal, with the rightmost edge of the bars indicating the central arrival latency. The hatched portion of the bars, located at the end of the signal with the earliest arrival times, represent a threshold, the length of which may vary based on modality of the first arriving signal. If the signal with the longer latency arrives within this threshold, simultaneity is perceived (top), whereas if it arrives after this threshold, temporal order is perceived. Figure adapted from Ulrich et al. (1987).*

An alternative model, named the 'temporal profile' model, can account for the possibility of perceiving asynchrony, without being able to discern temporal order (Stelmach & Herdman, 1991). It proposes that the visual and auditory signals feed into a single decision component, after being processed in two separate comparator components: a simultaneity comparator and a temporal order comparator. The temporal order comparator calculates the difference of its inputs, and the simultaneity comparator calculates the signal overlap. The decision component only considers the temporal order comparator when the TOJ task is being completed, and the simultaneity comparator when the SJ task is being completed.

This model predicts that it is possible for a small but noticeable SOA to produce a small output in the simultaneity comparator (due to the lack of temporal overlap), but a temporal order comparator output that does not vary enough from baseline to discern temporal order.

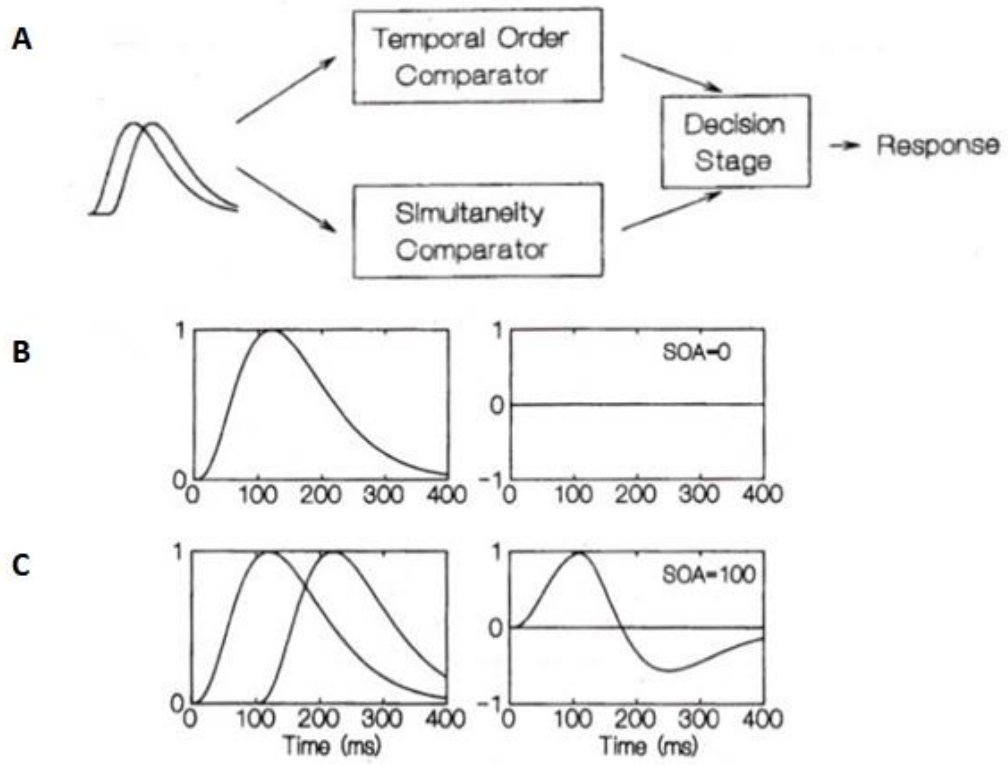


*Figure 2.3 - The temporal profile model. A) A schematic diagram of the model highlighting the auditory and visual signals (left), the two comparators and the decision stage. B) The auditory and visual signals (which, in this example, are entirely overlapping, see left subplot) and the resultant (hypothetical) signal in the temporal order comparator (right subplot). This calculates the difference between the inputs, leading to an output of "0" for physically simultaneous stimulil. C) The output of the temporal order comparator (right) in response to physically offset (SOA = 100) stimuli (left). Figure adapted from* Stelmach and Herdman (1991).

If the commonly reported lack of correlation between the PSS estimates of the SJ and TOJ tasks suggests the tasks are served by separate neural mechanisms (e.g., Love et al., 2013), then perhaps the two-stage model from Jaśkowski (1991) is more appropriate. This proposes two distinct 'centres' for simultaneity and temporal order judgements, which independently assess the difference in arrival latencies of audio-visual signals. The first signal to reach the simultaneity centre triggers a time interval, and simultaneity is perceived if the second signal reaches this centre within this period. If the stimuli are perceived as non-simultaneous then an 'order centre' output is considered. Such a finding is in line with fMRI research (Binder, 2015; Miyazaki et al., 2016), though see section 2.6.

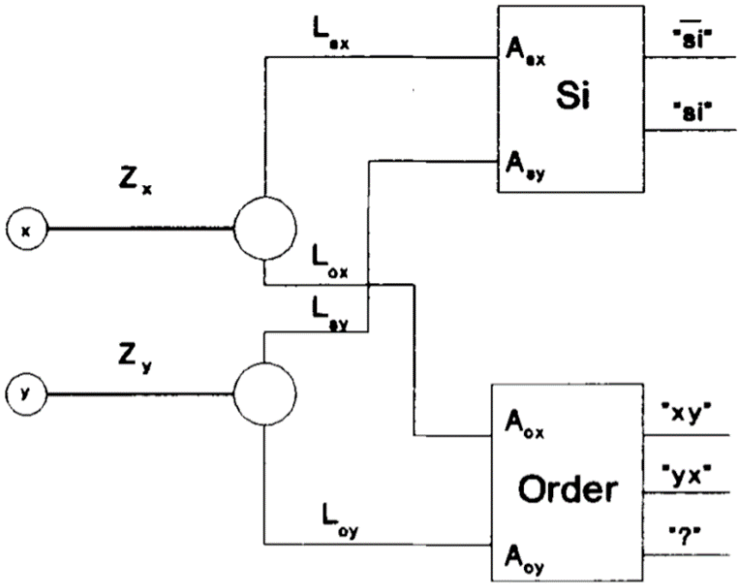


*Figure 2.4 - The two-stage model for order discrimination. Following a delay (Zx, Zy), the auditory and visual signals reach two points in the brain, which then send signals to the simultaneity and order centres at certain latencies ($L_{sx}$, $L_{sy}$, $L_{ox}$, $L_{oy}$). If the signals reach the simultaneity centre within a certain interval, simultaneity (si) is perceived. If non-simultaneity is perceived then the perceptual state generated in the order centre is considered, which will lead to either the perception of temporal order (xy or yx), or an inability to perceive order (?), while simultaneously recognising that the stimuli were not synchronous. Figure adapted from Jaśkowski (1991).*

Finally, the 'multisensory correlation detector' model (Parise & Ernst, 2016) provides a comprehensive account of the processes underlying judgements of simultaneity and temporal order (see section 3.7 for figures and model outputs). The MCD model is based upon the Hassenstein-Reichardt detector model of motion perception (Hassenstein & Reichardt, 1958), whereby two symmetrical subunits compare inputs from two adjacent receptive fields after a delay is applied to one of the inputs, and the resultant difference in the signals of the subunits is used to detect the direction of motion. Likewise, The MCD model suggests that the perception of simultaneity and temporal order both rely on two 'detectors' which compare the signals of two subunits. The outputs of the subunits are driven by the difference in the central arrival times of the auditory and visual stimuli, after a delay is applied to one of the two modalities. Akin to the two comparators of the temporal profile model (Stelmach & Herdman, 1991), the 'correlation detector' computes the temporal overlap of the two signals, whereas the 'lag detector' computes the temporal difference (see section 3.8 for these model outputs over varying SOAs). The lack of correlation commonly reported 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 (see figure 3.11). This model therefore proposes an alternative to prior claims of the existence of separate neural mechanisms underlying the two tasks, and provides a parsimonious explanation for uncorrelated estimates previously between the tasks. One notable drawback of the model, however, is that it cannot account for the known effect stimulus intensity has on processing latency (Carrillo-de-la-Pena et al., 1999) and the PSS (e.g., Leone & McCourt, 2015).

To summarise, only the two-stage model for order discrimination (Jaśkowski, 1991) proposes that judgements of simultaneity and temporal order are made in physically separate neural "centres". The threshold model (Ulrich, 1987), temporal profile model (Stelmach & Herdman, 1991) and MCD model (Parise & Ernst, 2016) all propose that the

decisional stage of these judgements relies on some form of shared process. The threshold model is the most simplistic, which proposes that the process is entirely shared, and TOJs are simply made when simultaneity is not perceived, though this model cannot account for the uncorrelated PSSs reported commonly when estimated using the two tasks (e.g., Van Eijk et al., 2008). The MCD and the temporal profile model are both similar, in that they propose that identical signals reach two separate subunits/comparators which compare the signals to assess either their temporal overlap or difference, the outputs of which feed into a shared decisional component. However, only the MCD model predicts that the outputs from both subunits are considered at the decisional stage of *both* tasks. Importantly, it is also only the MCD model that makes quantifiable predictions.

### 2.8. Reaction times as a measure of relative latencies

Simple reaction time (RT) tasks are a commonly used estimate of detection times. Simple RT tasks typically involve the presentation of a stimulus, which an observer must respond to (usually with a button press) as soon as they have perceived it. Importantly for this thesis, reaction times (RTs) have been argued to reflect processing latencies in the visual system (Pins & Bonnet, 1997). Such an argument is evidenced by the fact that RTs are correlated with neural processing latencies measured using EEG (Hülsdünker, Ostermann, & Mierau, 2019; Schomaker, 2009). However, it should be noted that reaction times do not represent absolute processing latencies, since RT responses occur after the neural signal reaches a decisional criterion (Cao, Zele, & Pokorny, 2007; Watson, 1986), the mean and standard deviation of which can vary based on individual differences and experimental variables (Grice, 1972).

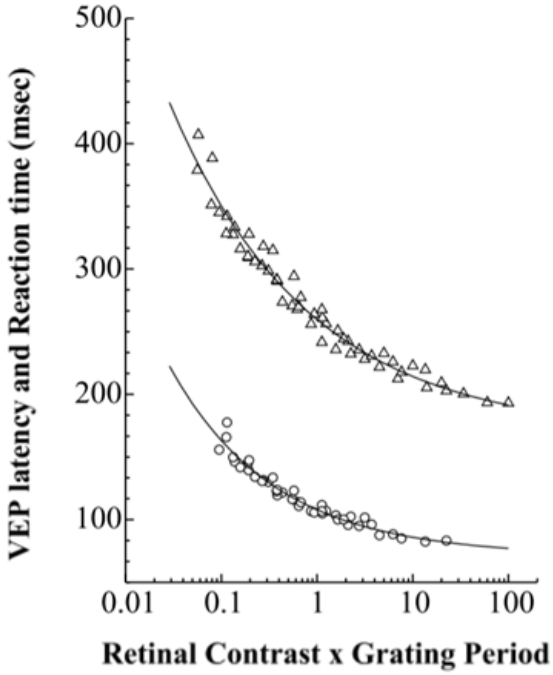


*Figure 2.5 - Simple reaction times (top curve) and visual evoked potential latencies (bottom curve), recorded at the Oz electrode of a single observer, as a function of retinal contrast multiplied by grating period. The additional time required for RTs, in comparison to the VEP latency, is evident. Figure taken from* Vassilev, Mihaylova and Bonnet (2002).

Akin to the commonly reported difference between the SJ and TOJ tasks, research has also highlighted a similar disparity between the simple RT and the SJ and TOJ tasks. Of course, there are various important differences between the simple RT task in comparison to the SJ and TOJ tasks, most notably being the single response button usually employed in RT tasks, and the reduced cognitive demand associated with this, in comparison to the two used for SJ and TOJ tasks. There is also a notable difference in both response strategies and speeds, where in the RT task observers are asked to respond as quickly as possible after perceiving the stimulus, in comparison to the SJ and TOJ tasks, where observers are usually asked to respond as quickly as possible after *making a decision* about the stimulus they have just perceived.

Nevertheless, comparable estimates of relative perceptual latencies can be obtained using the SJ, TOJ and simple RT tasks. For example, by calculating the difference in the RTs to unimodal audio and visual trials, it is possible to predict the SOA at which they must be presented, in order for the two stimuli to be detected simultaneously (Barnett-Cowan & Harris, 2009), an estimate that is comparable to the PSS. Additionally, Mégevand, Molholm, Nayak and Foxe (2013), reasoned that if the race model is violated then the two stimuli have been integrated into a singular percept, which they argued represented the range of SOAs where temporal order cannot be reliably discerned. They referred to these SOAs as an observer's temporal window of integration, an estimate which is reportedly comparable to the TBW measured using the SJ task. They discovered that the range of SOAs at which temporal order cannot be discerned is smaller than the range of SOAs at which the race model is violated. Importantly however, they argued that these results are indicative of a window of integration which varies, based on maximal performance at the task at hand/task demands, rather than reflecting mechanistically separate windows of integration that are utilised for each task.

Despite the clear differences in the methodology of the tasks, RT data can potentially be informative on the processes underlying relative timing judgement (SJ and TOJ) tasks. The aforementioned experiment by Barnett-Cowan and Harris (2009) showed a significant difference between the PSSs (of both the SJ and TOJ tasks) and the analogous predicted PSS (from the simple RT data). Such a finding is not necessarily surprising and may be driven by several factors. If we assume that SJs and TOJs are made by comparing the relative arrival time of two stimuli at multisensory areas of the brain to task-specific decisional criteria (Yarrow, Jahn, Durant, & Arnold, 2011), and also consider the decisional criteria which a signal is compared to when detecting a stimulus in an RT task (Henderson, 1970), then the raw estimates themselves may not be particularly informative, as any differences may be simply driven by the differential placements of the decisional criteria used to respond to

the same signal. However, if we assume that these decisional criteria remain somewhat consistent across different stimulus properties (though there is evidence that they *can* vary, see Yarrow, Jahn, Durant, & Arnold, 2011), then by varying said stimulus properties, such as stimulus intensity, and investigating whether the tasks have a shared dependence on this intensity manipulation, it would be possible to provide insight into whether these tasks share low level processing stages.

### 2.9. The effect of stimulus intensity on the PSS and RTs

If the difference in processing latencies of audio and visual stimuli are used to make judgements on perceived timing (e.g., Parise & Ernst, 2016; Stelmach & Herdman, 1991), then it could be expected that intensity driven differences in detection time could account for the corresponding effect of intensity on PSS (Roufs, 1963). While Roufs (1963) reported a highly similar visual intensity-dependent shift in the PSS and visual reaction times (between 30ms and 40ms), other researchers have reported intensity dependent changes in visual reaction times to be much greater than corresponding shifts in perceived simultaneity (Jaśkowski, 1992; Menendez & Lit, 1983). Thus, based on previous evidence, it is currently unclear whether the effect of intensity on processing latencies can explain the reported intensity dependent PSS shift.

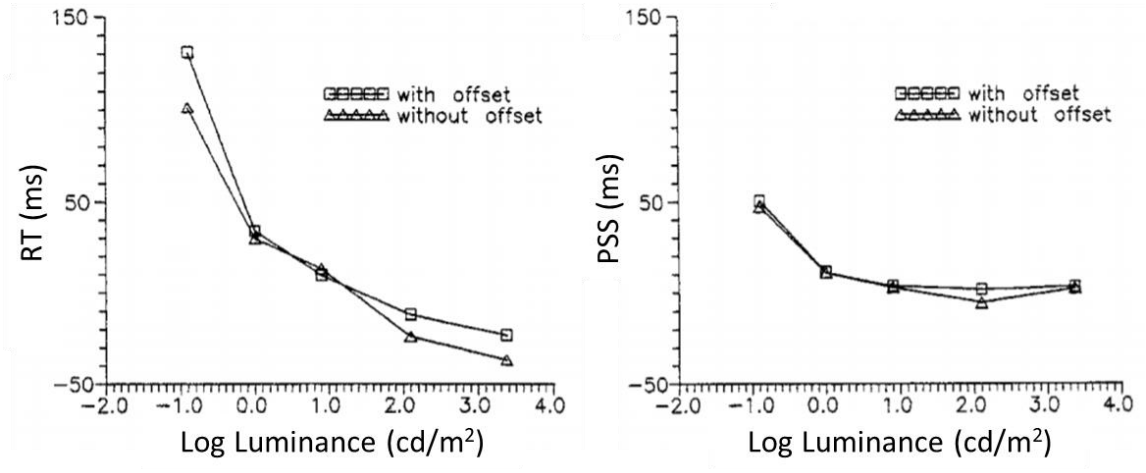


*Figure 2.6 - Simple RTs (left) for a unimodal visual stimulus, and PSSs (right) estimated via the method of adjustment (see Roufs, 1963), both as a function of log flash luminance in two separate stimulus presentation conditions. Figure adapted from Jaśkowski (1992).*

The reason for the discrepancy in the effect of intensity and RTs is disputed, and several theories have been proposed. These are discussed below:

One possible explanation for the previously reported discrepancy is that the effect of intensity on PSS may be reduced by observers focusing on the lower intensity stimulus of a stimulus pair in timing judgement tasks. This attentional shift would cause an increase in the processing latency of the high intensity stimulus (Di Russo & Spinelli, 1999), thereby reducing the difference in processing latency between two signals that would be caused by intensity. Such an attentional effect would likely not be evident in simple RT tasks (Jaśkowski, 1992). Alternatively, Sternberg and Knoll (1973) argued that both TOJs and RTs are based on the same internal response, but that the disparity could be driven by a different utilisation of this response. They argue that the focus on speed for RTs, means that a response is made after the signal passes a detection threshold, whereas the focus on prevision in TOJs means that the judgement is made when the signal is at its peak, and changes in intensity may impact the time at which the signal reaches the detection criterion and peak, differentially (Sternberg & Knoll, 1973). Another possibility lies in the

use of different stimulus intensities across blocks in the SJ and TOJ tasks (e.g., Kopinska & Harris, 2004; Menendez & Lit, 1983; Roufs, 1963). Presenting individuals with a biased distribution of stimulus offsets can shift perceived simultaneity away from the peak of the stimulus distribution (Miyazaki, Yamamoto, Uchida, & Kitazawa, 2006). Since less intense stimuli are perceived later (Roufs, 1963) it is conceivable that the average perceived stimulus offset will differ across blocks with different relative visual intensities, thus having divergent impacts on perceived simultaneity, and therefore obfuscating a potentially larger PSS shift. It is so far unclear as to whether this disparity in the intensity dependency of simple RTs and the PSS will be eliminated if various visual intensities were presented within the same block. Furthermore, it is currently unknown as to whether this intensity dependency is shared across the audio-visual SJ and TOJ tasks.

Cardoso-Leite, Gorea and Mamassian (2007) presented a pertinent theory on what causes the discrepancy in findings between the RT and the TOJ task. They argued that a 'one-system-two-decisions' approach, where different decision processes operate serially on the same internal signal, could explain the previously reported differences. By classifying RTs based on the TOJ decisions made immediately after them (i.e. classifying the reaction times based on which stimulus was seen first), they could predict reaction times based on these decisions. The results highlighted that RTs were slower to stimuli when the stimulus which came physically second was perceived first. It was therefore concluded that the perceptual decisions are based on the same internal signal as the motor decisions (i.e. RTs), but were made relatively later and with different decisional criteria.

### 2.10. Proposed separate subsystems underlying 'action' and 'perception' tasks

While unimodal TOJs and RTs may be based on the same internal signal (Cardoso-Leite, Gorea & Mamassian, 2007), the mechanisms involved in the underlying processing may differ between the tasks, dependent on whether the bimodal stimuli are presented simultaneously, or with a physical offset. Leone and McCourt (2009) ran several

experiments with audio-visual stimuli at varying stimulus onset asynchronies and intensities. They reported evidence of multisensory enhancement (i.e. race model violations) when the bimodal stimuli were presented simultaneously, regardless of stimulus intensity. This interesting finding may seem counterintuitive, as one might expect that increasing the intensity of one stimulus in a simultaneously presented bimodal pair, would reduce its processing latency, causing a difference in the arrival times of the signals at multisensory neural regions and potentially mitigating the resultant multisensory enhancement. In a further experiment, Leone and McCourt (2015) ran the SJ, TOJ and simple RT tasks under a scotopic and photopic light condition. They found that changing the brightness of the visual stimulus shifted the observers' PSSs in the SJ and TOJ task. Yet, in line with Leone and McCourt (2013) (and in contrast to Mégevand et al., 2013), race model violations were only evident when the stimuli were presented simultaneously. They concluded that the multisensory processing required to complete the two tasks is mediated by two separate subsystems. These consist of a 'perception' subsystem for the SJ and TOJ tasks, and an 'action' subsystem for the RT task. It was argued that coincident detectors tuned to specific stimulus intensities, are only activated when stimuli of specific intensities are presented at the correct arrival latency differential to have been produced by *physically* simultaneous stimuli. Thus, these coincident detectors are proposed to be able to compensate for intensity-dependent processing latencies. Importantly, they argued that this process only occurs in the 'action' (RT) task, and not the 'perception' (SJ and TOJ) tasks. The reasons for why intensity-dependent processing latencies are also not compensated for during the SJ and TOJ tasks, however, are unclear (see section 4.4.2).

The findings of Leone and McCourt were further supported by Harrar, Harris and Spence (2017) who recalibrated the observers' PSSs through an adaptation period, where repetitive audio-leading, visual-leading or simultaneous stimuli were presented (see section 2.11). They found that regardless of the pre-test adaptation phase (and consequently the

observers' PSSs), violations of the race model for multisensory RTs were only evident at an SOA of 0ms, suggesting that some form of temporal compensation occurs when performing bimodal audio-visual simple RTs (Leone & McCourt, 2015).

### 2.11. The recalibration of perceived simultaneity

Our physical environment is constantly changing, and therefore so too is our audio-visual perceptual experience. As a person moves through a physical space, the relative central arrival times of the two modalities fluctuate due to stimulus properties, and the perceptual system must account for this variability to accurately perceive audio-visual simultaneity. Two potential strategies which could be followed (assuming that coincident detectors do not compensate for differences in central arrival latencies when making perceptual judgements, see Leone & McCourt, 2015), in order to account for the varying arrival times of two (at least partially) independently processed signals, with different modality and stimulus specific processing delays, are the following: 1) perceptual timing mechanisms could be calibrated to capture long term average relative signal delays and/or 2) a mechanism could continually adjust the perceptual system to compensate for these varying latencies. Recent research has provided evidence of the latter.

Congruent with evidence that past sensory experiences of audio-visual events are used as priors to decide which signals have emanated from the same source (Habets et al., 2017), the PSS has been shown to flexibly shift, based on prior experience. This was first examined by Fujisaki, Shimojo, Kashino and Nishida (2004), who presented repetitive temporally offset flash-bleep 'adaptation' stimuli to participants for 3 minutes. This adaptation stage was followed by simultaneity judgement trials, each preceded by a 10 second 'top up' of additional temporally offset stimuli. It was found that an individual's PSS, estimated using the SJ trials, would shift towards the leading modality of the adaptation trials. Furthermore, larger temporal offsets in the adaptation trials lead to larger shifts in perceived

simultaneity. In other words, the brain recalibrates what it perceives as simultaneous based on recent sensory experience.

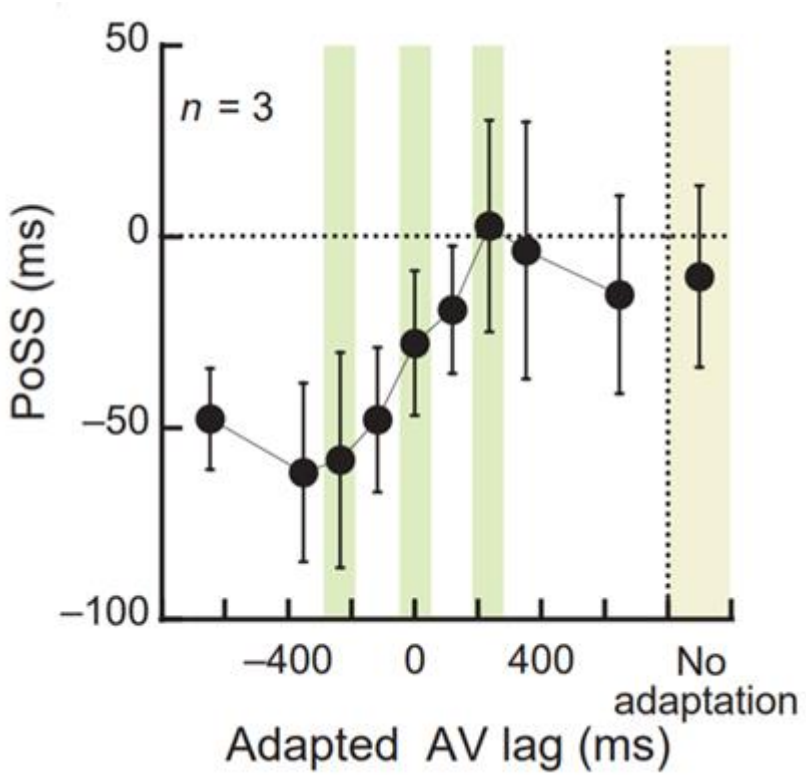


*Figure 2.7 - The recalibration of the point of subjective simultaneity (here abbreviated as PoSS) as a function of the adapted audio-visual lag (i.e., the physical SOA of the* adaptation *trials). The right-most column reflects the PoSS when estimated following no adaptation trials. Figure adapted from Fujisaki, Shimojo, Kashino and Nishida (2004).*

The recalibration of perceived simultaneity appears to be a robust finding. The same PSS shift is also evident when estimated using the TOJ task (Vroomen, Keetels, De Gelder, & Bertelson, 2004). Again, the PSS shift was in the direction of the leading modality of the adaptation trials, and larger temporal offsets (up to 200ms audio or visual-leading) in these trials led to larger shifts in PSS. Fujisaki et al. (2004) expanded on their novel finding, by showing that the adaptation effect was also evident for more complex stimuli, consisting of the stream/bounce illusion, where observers decided whether two balls, with an accompanying 'collision' sound (with varying offsets), passed through or bounced off each

other. Interestingly, adaptation trials consisting of flash-beep stimuli led to changes in perceived simultaneity of stream/bounce stimuli. The authors highlight that changes in early relative processing latencies of the two modalities could underlie such an effect. However, through re-analysing Fujisaki et al.'s (2004) data set, Yarrow et al. (2011) argued that changes in decisional criteria, as opposed to changes in early processing, provide a more parsimonious explanation of the data (Yarrow et al., 2011).

The effect of recalibration is not restricted to audio-visual perception, nor is it restricted to the modality pairing which is presented to observers through adaptation trials: Di Luca, Machulla and Ernst (2009) replicated the effect of audio-visual perceptual recalibration, but also showed that the shift in perceived simultaneity from audio-visual recalibration trials can transfer to shifts in the relative timing perception of other modality pairings. For example, following visual-leading audio-visual adaptation trials, the PSS shifted towards the visual-leading SOAs of a visual-*tactile* stimulus. The authors argued that this cross-modal transfer of perceptual recalibration represents a reduction in the perceptual latency of the visual stream, to compensate for the previously experienced audio lag, a claim that was bolstered by a follow-up simple unimodal RT task. The RT task revealed that visual RTs were significantly faster following audio-leading adaptation trials, than when following visual leading adaptation trials. Such a finding underlines the possibility that recalibration is driven by changes to relative processing latencies, rather than shifts in the decisional criteria which these relative latencies are compared against. It should be noted that this effect was only apparent when the audio stimuli were presented through a speaker that was co-localised with visual stimulus, as opposed to presented separately via headphones. The authors argue that this finding suggests that the perceptual system varies the processing latency of the sensory estimate which has the lowest relative trust (i.e. the relative trust in the perceived timing of an auditory stimulus is lowered when it is presented from a spatially separate location than a visual stimulus, such as through

headphones). Complementary research has shown that visual-leading recalibration trials reduced the simple reaction times to unimodal auditory stimuli, whereas auditory-leading recalibration trials increased reaction times (Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009). Again, the contrast in which modality sees a RT shift between the two experiments can perhaps be attributed to which modality the observer has the lowest relative trust in (Di Luca et al., 2009).

Interestingly, more recent research has shown that a significant shift in PSS could only be induced when the observer was aware of the temporal offset in the adaptation trials (Tsujita & Ichikawa, 2016), suggesting the involvement of higher-order processes in this perceptual shift. However, it must be noted that this awareness was produced both through presenting trials with a noticeable temporal lag from the very first trial (as opposed to gradually increasing the offset over time), and by *informing* the observer of its presence. This process of informing observers of the lag is important, since even slight variations in task instructions have been shown to affect the placement of decisional criteria (Yarrow & Roseboom, 2017), and the possibility of this inducing a PSS shift, that would perhaps be otherwise undetectable due to the small sample size, cannot be excluded. Additionally, these findings contrast with prior research where a PSS shift was observed, in tasks which aimed to minimise awareness of the temporal offset in adaptation trials by gradually increasing this over the initial period of adaptation (Di Luca et al., 2009; Machulla, Di Luca, Froehlich, & Ernst, 2012; Van Der Burg, Alais, & Cass, 2015).

#### 2.11.1. Rapid recalibration

The effect of recalibration on perceived simultaneity is not a permanent one; the shift in PSS due to perceptual recalibration declines as the number of test trials since the adaptation phase increases (Machulla et al., 2012). This gradual reduction was found to be due to the presentation (and therefore perception) of stimuli with a leading modality or offset that is inconsistent with that of the adaptation trials. This result suggests the brain

continually adapts its perception of simultaneity based on its most recent evidence (Machulla et al., 2012). This theory was further expanded by Van der Burg, Alais and Cass (2013), who found that perceived simultaneity can shift, based on the leading modality of just a *single* preceding trial. Van der Burg et al. (2013) also highlighted that trial n-2 had a significant, though noticeably smaller effect than the immediately preceding trial (n-1), suggesting that a single trial does not entirely overwrite the timing 'evidence' of other previous trials. Again, the effect of this *rapid* recalibration is evident for more complex stimuli (Van der Burg & Goodbourn, 2015) along with simple flash-beep trials (Van der Burg et al., 2013).

Interestingly, research has shown that both prolonged audio-visual recalibration (i.e. to repetitively presented adaptation trials) and rapid recalibration can be demonstrated simultaneously within a single experimental paradigm, and that these effects operate on different time frames (Van Der Burg et al., 2015). The PSS shift induced by prolonged recalibration gradually decays over time, whereas the rapid recalibration induced shift remains relatively constant regardless of the time passed since a previously experienced stimulus and is based solely on the leading modality of the immediately preceding trial. Related results have shown that the time courses of rapid recalibration and prolonged recalibration respond differently to training in the simultaneity judgement task (De Niear, Noel, & Wallace, 2016), and it is therefore possible that there are separate mechanisms underlying the two effects (Van der Burg et al., 2015). However, while there is evidence that recalibration to the prolonged presentation of adaptation trials is driven by changes in relative processing latencies (e.g., Di Luca et al., 2009), it is unknown whether the same mechanism underlies rapid recalibration.

Akin to the inter- and intra-individual differences often reported in the estimated TBW size of observers (Foss-Feig et al., 2010; Hairston et al., 2005; Hillock-Dunn & Wallace, 2012; Lee

& Noppeney, 2014; Scarpina et al., 2016; Stevenson et al., 2014), the PSS shift induced by the audio-visual offset of a preceding stimulus is highly variable. Cross-sectional research using flash-beep stimuli has shown that rapid recalibration shows a U-shaped developmental pattern, with a larger effect seen during both the teenage and latest years of life (Noel et al., 2016), corresponding with the developmental pattern of TBW size (Bedard & Barnett-Cowan, 2016; Hillock-Dunn & Wallace, 2012; Noel et al., 2016). Furthermore, the size of the PSS shift seen in rapid recalibration has been shown to be positively correlated with TBW size (Van der Burg, Alais, & Cass, 2013), an effect which has been replicated with both flash-beep (Noel et al., 2016; Simon, et al., 2017) and speech stimuli (Noel et al., 2016).

If rapid recalibration exists to compensate for differences in neural propagation times (Fujisaki et al., 2004), then it is perhaps unsurprising that those who perceive simultaneity over a wider range of audio-visual offsets will show larger PSS shifts as a result of rapid recalibration. What is surprising, however, is that rapid recalibration has been shown to be greatly diminished in those with ASD (Turi, Karaminis, Pellicano & Burr, 2016), a developmental disorder typically associated with increased TBW size (Foss-Feig et al., 2010; Stevenson et al., 2014), and to negatively correlate with autistic traits in non-autistic individuals (Turi et al., 2016). Such a finding may link to evidence from research into prolonged recalibration, which shows that while those who are musically trained have smaller TBWs, they do not show a smaller recalibration-dependent shift in PSS (Jicol et al., 2018). Furthermore, while this reduced PSS shift is evident for flash-beep (Turi et al., 2016) and tool stimuli (Noel, De Niear, Stevenson, Alais, & Wallace, 2017), the difference between typically developing and autistic individuals is non-existent for more complex speech stimuli (Noel et al., 2017). Thus, it is It is therefore apparent that while the size of the TBW and the PSS shift are associated, they must also rely on (at least partially) independent processes. Additionally, these processes are highly dependent on stimulus

properties and individual differences in observers, meaning that it is difficult to isolate the potential idiosyncrasies of each by comparing different pieces of research with separate samples.

While rapid recalibration has been consistently replicated using the SJ task, the TOJ task has so far produced much less conclusive findings (Ju, Orchard-Mills, Van Der Burg, & Alais, 2019; Keane, Bland, Matthews, Carroll, & Wallis, 2020; Recio, Cravo, de Camargo, & van Wassenhove, 2019; Roseboom, 2019). Interestingly, a study combining alternate TOJ and SJ trials provided evidence that the effect of rapid recalibration is dependent on the physically leading modality of the previous trial, rather than the *perceived* leading modality (Van der Burg, Alais, & Cass, 2018). Analogous to the dependency of facilitative multisensory integration on the physical, rather than the perceived temporal offset of audio-visual trials (e.g., Harrar et al., 2016; Leone & McCourt, 2015), the authors argued that information of the true, physical offset between stimuli may be available to the observer, despite the commonly reported discrepancy between physical and perceived simultaneity. Furthermore, the authors posited that rapid recalibration must therefore be driven by an early timing mechanism, which an individual has no conscious access to (Van der Burg et al., 2018). However, the potential mechanism that could underpin this proposed process was unspecified.

In contrast to research from Van der Burg et al. (2018), relevant EEG findings provide evidence that rapid recalibration is driven by decisional processes, rather than an early timing mechanism. The results highlighted that the leading modality of a preceding trial modulates the late components of the ERP in central and parietal electrodes (Simon, Noel, Wallace, et al., 2017), corresponding with the late ERP components involved in perceptual decision making (Kelly & O'Connell, 2013; O'Connell, Dockree, & Kelly, 2012; Tagliabue et al., 2019).

Research involving more ecologically valid audio-visual speech stimuli also reported significant differences in the ERP, dependent on the leading modality of the preceding stimulus (Simon, Nidiffer, & Wallace, 2018). These differences were centred mainly on centro-parietal regions and were reported between 280ms and extended 804ms. Furthermore, it was reported that there was a significant correlation between the difference in ERP amplitude between pooled auditory leading and visual leading SOAs (i.e. [450ms + 300ms, + 150ms AV] - [150ms + 300ms + 450ms VA]), and the SOA of the previous trial. The fact that the difference in ERP amplitude increased at larger SOAs mimics behavioural evidence that the effect of rapid recalibration is greatest following larger offsets at trial n-1 (e.g., Van der Burg et al., 2013).

The evidence in favour of the existence of some form of recalibration mechanism(s) appears to be unanimous. However, the ecological benefit of this rapid recalibration mechanism is not currently known, as the physical temporal offsets of real-world stimuli do not typically vary rapidly (e.g., Van der Burg et al., 2018). One potential reason for such a mechanism to exist, would be to compensate for rapidly varying central arrival times related to changes in stimulus intensity, however this possibility was previously untested.

### 2.12. Perceptual learning

Perceptual learning is a term that refers to a wide array of experience-dependent improvements in the processing of sensory information (Gold & Watanabe, 2010). These improvements can be brought on through 'every day' sensory experience, to the perceptual learning induced through psychophysical paradigms (Seitz, 2017). Such changes include improvements in spatial frequency (Fiorentini & Berardi, 1980) and texture discriminations (Karni & Sagi, 1991), to more high-level extraction of complex relationships between stimuli, such as that between pieces on a chess board (for a review, see Kellman & Garrigan, 2009). As well as structural changes, such as those seen in the sensory cortices (e.g., Fahle & Poggio, 2002), perceptual learning can also be driven by changes in decisional

criteria (Aberg & Herzog, 2012; Herzog & Fahle, 1999), for example, the expectation of a visual stimulus has been shown to enhance its detection rate through an altered decision criterion (Bang & Rahnev, 2017). Importantly, the effect of perceptual learning on decisional criteria and perceptual sensitivity can be differentiated (Aberg & Herzog, 2012).

A common theme found across perceptual learning research is the variability of the generalisability of perceptual improvements across stimuli with different features. For example, in certain paradigms, training induced improvements are reportedly not generalisable across stimuli with different spatial frequencies and orientations (Fiorentini & Berardi, 1980) or direction or speeds of movement (Green, Kattner, Siegel, Kersten, & Schrater, 2015). Yet there is evidence of learning transfer across some perceptual tasks (Green et al., 2015; McGovern, Roach, & Webb, 2012) and across variations in the pattern of visual stimuli (Green et al., 2015). There are various, task-dependent reasons which may explain this variability in generalisability, as generalisability is typically reduced for more difficult tasks, and for perceptual learning relating to low-level stimulus features (Fahle, 2005). Furthermore, generalisability has been shown to be reduced following higher amounts of training (Jeter, Dosher, Liu, & Lu, 2010).

### 2.12.1. Training to reduce size of the temporal binding window

As discussed previously, individuals with certain disorders, such as autism (Foss-Feig et al., 2010; Stevenson et al., 2014) and schizophrenia (Stevenson et al., 2017), typically have larger than average TBWs. Importantly, deficiencies in multisensory processing predict higher susceptibility to the McGurk illusion (Stevenson, Zemtsov, et al., 2012), and poorer performance on speech perception tasks (Conrey & Pisoni, 2006). This may be a result of a general deficiency early in the multisensory processing stream, as those with ASD do not show the N1 amplitude reduction that is observed for typically developing individuals following the presentation of a simultaneous audio-visual stimulus (Brandwein et al., 2013). Critically, increased TBW size has been shown to cascade to impairments in speech

perception (Stevenson et al., 2018), and has been associated with increased fall risk in older age (Setti et al., 2011).

A recent focus of perceptual learning research has been on training participants to reduce the size of their TBW (e.g., Zerr et al., 2019). It is feasible that by training observers to improve the accuracy of their simultaneity judgements, and reduce the size of their TBW, that they will experience notable everyday improvements. Over five days, Powers III, Hillock and Wallace (2009) provided feedback following responses in either a two-alternative forced choice or two-interval forced choice SJ task. They reported a significant reduction in TBW size after just a single training phase in both task variants. Importantly, the reductions in TBW size seen after day 5 of training remained stable at a one-week follow-up, suggesting that training in this manner could lead to long term improvements, though no longitudinal studies have been conducted to confirm this. The promising findings from Powers III et al. (2009) have led to the proposal from several researchers for TBW training to be used as an intervention for those with multisensory deficiencies (Foss-Feig et al., 2010; Powers et al., 2009; Setti et al., 2014; Stevenson et al., 2017; Wallace & Stevenson, 2014).

For both variants of the SJ task used by Powers III et al., the reduction in TBW size was only significant for light-first SOAs, which is not necessarily unexpected given the asymmetry of the TBW (see section 2.3). The reduction in temporal binding window following bimodal SJ training has been replicated several times, though other researchers have reported that the reduction in TBW size only occurs for *audio*-leading SOAs (De Niear et al., 2016), whereas Stevenson et al. (2013) found that both the audio-leading *and* visual-leading TBWs can be reduced through training. Again, inconsistencies are not particularly surprising, given that the TBW can be trained independently for audio-leading and for visual-leading

trials, with no generalisability of training-induced improvements between trials with the opposite leading modality (Cecere et al., 2016).

The reason for why there is such a discrepancy in TBW training data, in terms of which leading modality sees improvement, is currently unclear. One possibility lies in individual differences between samples, where different observers have differing audio-leading and visual-leading TBW sizes. Since the reduction in TBW size is dependent on the SOAs used in the training paradigm being suitably small for each observer (De Niear et al., 2016), it is possible that the SOAs used in the training paradigms of different experiments were more suited to elicit a reduction to a particular side of the TBW for that specific sample. Another possibility relates to the fact that these experiments did not attempt to match the intensity of their audio-visual stimuli. Comparable to how, following a temporal recalibration period, reaction times shift for the sensory estimate which has the lowest relative trust (Di Luca et al., 2009), perhaps the side of the TBW which reduces in size following training depends on which modality is the most reliably perceived (i.e. intense). What is clear, however, is that observers must be trained with both audio-leading and visual leading trials, ideally at SOAs tailored to each individual, in order for training to be maximally effective.

The mechanisms behind how the TBW is reduced through training are not fully understood. Matthews, Welch, Achtman, Fenton and FitzGerald (2016) trained observers with either the SJ or TOJ task. They reported improvements in performance following training, but importantly, argued that due to the distinct reaction time patterns seen across SOAs for the two tasks, that that these improvements were driven by decisional factors, rather than changes in perceptual sensitivity. Such a conclusion aligns closely with work which shows that the size of the TBW can be manipulated by varying the instructions given to participants (Yarrow & Roseboom, 2017). The size of the TBW is highly dependent on the placement of decisional criteria (Yarrow et al., 2011), and simply asking observers to be

more conservative when making SJs led to a reduction in TBW size (Yarrow & Roseboom, 2017). It is therefore entirely possible that the training-induced reductions in TBW size observed by previous researchers (e.g., Powers III et al., 2009), were due, at least partially, to shifts in decisional criteria.

In contrast to the proposed influence of decisional processes, neuroimaging research suggests an influence of early sensory processing mechanisms on training induced improvements in judgements of audio-visual simultaneity and temporal order. Bernasconi, Grivel, Murray and Spierer (2010) found that unisensory auditory TOJ training led to changes as early as 43ms in the AEP. This was furthered by Powers, Hevey and Wallace (2012), who demonstrated that reductions in TBW size are associated with a wide range of changes in both early and later processing stages. These included the reduced BOLD signal in primary auditory and visual sensory cortices, proposedly indicative of a shift in activity to a more specialised selection of neurons following training (Powers III et al., 2012). Furthermore, they reported reduced BOLD signal in the posterior superior temporal sulcus (pSTS), an area commonly associated with the processing of complex perceptual information (Allison, Puce, & McCarthy, 2000) and the perception of audio-visual simultaneity in particular (e.g., Calvert, Hansen, Iversen, & Brammer, 2001). Importantly, they discovered an increase in the functional coupling between the pSTS and the primary auditory cortex following training, suggesting that the reduction in TBW also results from changes in feedback mechanisms from the pSTS to the auditory cortex.

Consistent with the unpredictable generalisability of training induced improvements in other types of perceptual learning research, training observers using audio-visual perceptual judgement tasks leads to varied levels of transferability between tasks and stimuli. Matthews et al. (2016) showed that training in the SJ task led to no improvements in TOJ performance, and TOJ training led to no significant SJ improvements. In contrast,

unisensory visual TOJ training has been shown to reduce TBW size measured using the audio-visual SJ task (Stevenson et al., 2013). Likewise, audio-visual TOJ training leads to a reduction in susceptibility to the sound-induced flash illusion (Setti et al., 2014), which is another measure of audio-visual temporal acuity (McGovern et al., 2014), yet audio-visual *SJ* training leads to no reduction susceptibility (Powers, Hillock-Dunn, & Wallace, 2016).

It is so far uncertain as to the specific reasons behind the inconsistency in the generalisability of learning with the aforementioned tasks, though this could relate to task difficulty (Fahle, 2005), the stimuli used (Fahle, 2005) or the amount of training carried out (Jeter et al., 2010). While the real-world importance of the generalisability of training-induced improvements across different perceptual tasks, or across different stimulus modalities, is uncertain, the importance of the generalisability of training across stimulus *properties,* such as intensity, is clear. Since relative stimulus intensities perpetually fluctuate in an individual's immediate environment, it is vital that any training-induced improvements can be generalisable across differing intensities. Without such generalisability, training-induced improvements in audio-visual perception, in those with multisensory deficiencies, would not be beneficial across a practical range of real-world stimuli.

### 2.13. Limitations of past research and aims of the current thesis

While audio-visual timing perception has been studied extensively, there remained many open questions, particularly those which are tied to stimulus intensity, which this thesis aimed to address:

1) While it is known that processing latencies are intensity-dependent, it is unclear as to whether they alone can explain the intensity-dependency of perceived simultaneity, or whether this PSS shift is alternatively driven by higher-level

processing. **Chapter 5** aimed to investigate whether intensity-dependent response latencies can account for the previously reported effect of intensity on PSS.

2) The PSS estimated using the SJ and TOJ tasks have been commonly shown to be uncorrelated, an effect which **chapter 4** aimed to replicate. It is currently unclear whether this is due to differences in low-level processing stages, or whether this was driven by differences at the decisional stage. **Chapter 6** aimed to contribute to this discussion, by investigating the difference in the time course of the neurocognitive processes underlying the two tasks.

3) The multisensory correlation detector (MCD) is a compelling processing model, that can explain the uncorrelated PSS estimates of the SJ and TOJ tasks. However, this model does not account for the effect of stimulus intensity on the PSS. **Chapter 5** aimed to address this limitation and to extend the model to account for the stimulus intensity-dependent processing latencies.

4) Recent research shows that the PSS rapidly recalibrates based on the leading modality of a single trial. Since processing latencies are dependent on stimulus intensity, **chapter 6** aimed to investigate whether the PSS also rapidly recalibrates based on the intensity of the preceding trial. In other words, it aimed to investigate whether an intensity dependent processing delay, which leads to an offset in the central arrival times of audio-visual stimuli, would be rapidly recalibrated to as if it were a physical offset.

5) Research has shown that through training, the size of the TBW can be reduced, which could potentially lead to real-world improvements in those with multisensory deficiencies. However, in order for these improvements to be beneficial across the wide array of stimuli that are perceived in real-world environments, they must be generalisable across stimuli with different properties. Given that the generalisability of perceptual learning is inconsistent, depending on

the stimuli used in the training paradigm, **chapter 7** aimed to investigate whether reductions in TBW size can be generalised between stimuli of differing intensities.

## 2.14. Chapter Summary

This chapter covered the tasks most commonly used to assess audio-visual relative timing perception, and then summarised their wide array of applications, and the perceptual processes which have been studied through their use. This included the adaptable temporal window within which audio-visual stimuli are combined to form a singular percept and the recalibration that occurs during perceptual processing to account for rapidly changing audio-visual stimulus offsets. One common theme reported throughout this chapter was the difference in findings often reported by the SJ and TOJ tasks. The following section discusses the various methods used throughout this thesis and discusses the MCD model which accounts for the processes underlying these tasks in detail.

# 3. General methods

## 3.1. Introduction

The previous chapters summarised the basic neurological processes underlying audio-visual perception, and discussed in detail the various experiments which have been used to understand the array of processes underlying the perceived relative timing of audio-visual stimuli. The current chapter discusses some of the commonly utilised methods used to investigate these processes and discusses how these paradigms were utilised to best address the various research questions presented in this thesis. Further, it discusses a processing model used to explain the differences in the SJ and TOJ tasks, alongside a fitting method used to fit the data collected by the author. Initially however, it describes the equipment that was purpose built by the author to conduct these experiments, and covers the adaptive procedure used to determine the stimulus intensities used in several chapters of the presented thesis.

## 3.2. Equipment

The visual stimuli across chapters 5 to 7 were produced using a LED that was sealed inside of a manually modified plastic encasing. The LED was located behind a combination of neutral density filters to reduce the intensity of the light reaching the observer (see figure 3.1B). An electric saw was used to cut an opening into the face of this box, and a foam pad was super glued to the inside of the box and lid. A central groove was cut in the foam, leading from the back of this box to the open face, and a small hole was drilled through the back of the box, and the LED was placed into the open groove. Small slits were cut into foam located in both the box and lid to hold the neutral density filters in place (figure 3.1B).

A B

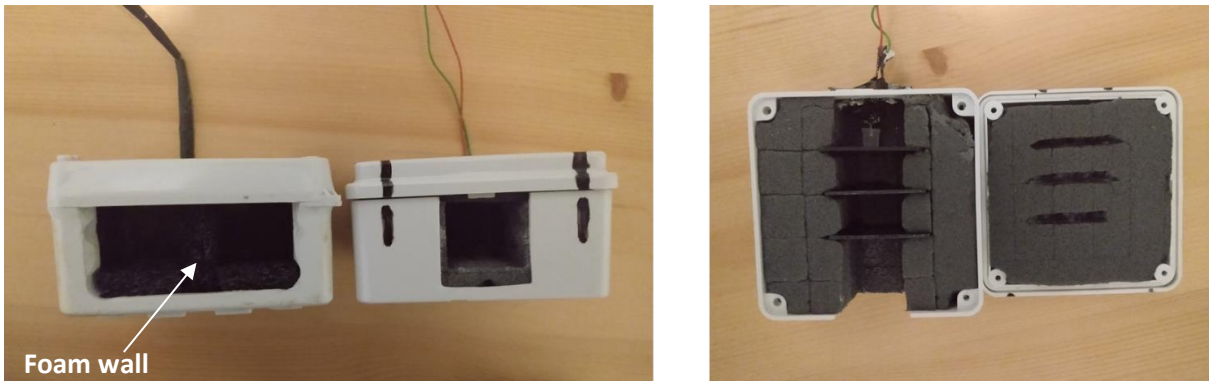


*Figure 3.1 A) The modified plastic encasings, one of which contains a single opening (A - right), whereas the other contains two openings, separate by a foam wall (A - left). B) A top-down view of the plastic encasing containing a single opening with the lid removed. Three neutral density filters are visible in front of the LED (B - left), as well as the slits in the lid to keep the filters in place (B - right)*

The absolute visual and auditory thresholds of observers were estimated for the purpose setting a range of supra-threshold stimuli for use in chapters 5 to 7. To briefly summarise, participants were presented with either two clearly visible flashes, one of which was followed by a low intensity auditory tone, or with two clearly audible tones, one of which was followed by a low intensity visual flash. For this preliminary experiment, a second plastic encasing was modified to contain two LEDS, each placed behind different sets of neutral density filters, separate by a foam wall (figure 3.1A, left). This allowed for the presentation of a high intensity visual stimulus through one set of filters, and a range of low intensity visual stimuli through the other set, both within the same experimental session, and importantly, from roughly the same location. Auditory stimuli across all experiments in chapters 5 to 7 were produced using a single 'Xenta M-219 Notebook speaker' situated 1.62° below the window which the LED was located behind.

A project box was modified for use as a button box for observers to provide responses (see figure 3.2). This contained two shallow buttons connected to a wire which was fed out through a small hole drilled into the face of the box. The use of shallow buttons (see figure 3.2 legend) are important, as minimising the movement required to make a response is vital in recording accurate RT data (Brenner & Smeets, 2019). Furthermore, it was important that this button box was small and manoeuvrable, to ensure participants could rotate this between blocks (see section 3.6).

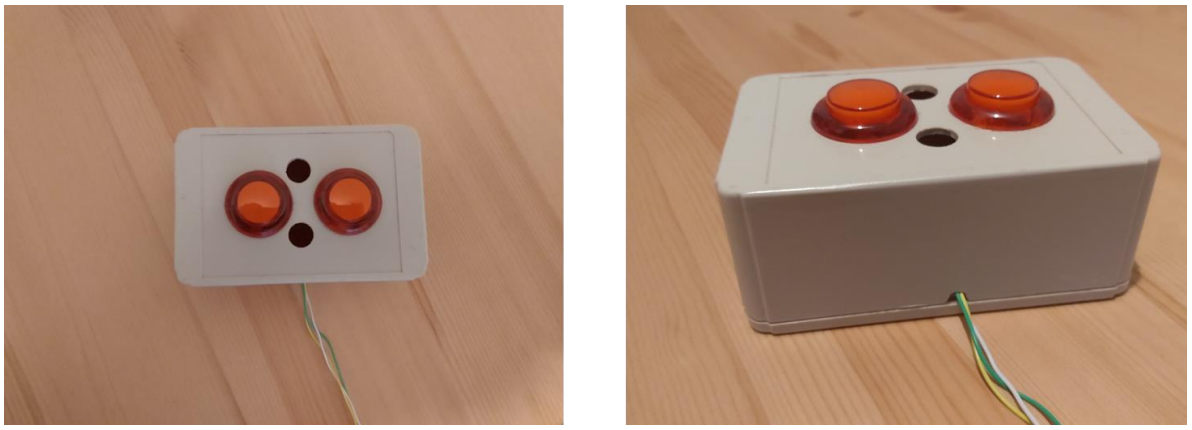

*Figure 3.2 – top down (left) and angled view (right) of the button box. The two unused holes pictured here initially contained buttons, however, these were considered to be too deep, in that there was a longer time between initiating a button press and this press being registered, relative to the shallow buttons pictured here. This additional time would have produced response times less reflective of the observer's actual detection times, in comparison to the shallow buttons.*

Finally, for the pilot experiment presented in chapter 4, a red LED and piezo sounder were soldered to an Arduino Genuino Micro. This was connected to a computer mouse with two buttons.

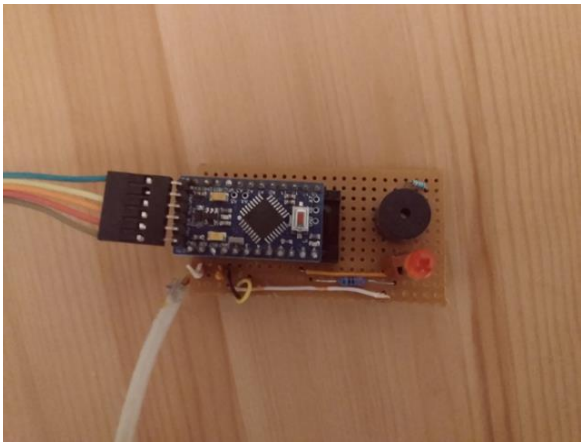

Figure 3.3 – Arduino Genuino Micro, with piezo sounder (top right of circuit board) and red LED (bottom right of circuit board). Computer mouse not pictured.

### 3.3. Psychophysical paradigms

Psychophysics is an experimental method often used alongside, or in place of neuroimaging methods, to quantify the relationship between a physical stimulus and its sensation. To quote Professor David Heeger (2006, para. 1), who highlighted the importance of psychophysical methods, by likening using only neuroimaging to understand the structure and activity of the brain, to viewing only the individual pieces of a car: “If you know only how those components work, you still do not know what the car is for”.

A wide range of psychophysical paradigms have been utilised to study a broad range of topics, such as luminance and contrast discrimination (Laming, 2013; Whittle, 1986), the perception of colour and material (Brainard, Cottaris, & Radonjić, 2018), the effect of heat on pain thresholds (Kojo & Pertovaara, 1987) and the effect of odour exposure on olfactory sensitivity (Dalton, 2000). Psychophysical paradigms typically involve presenting a varying stimulus (or stimuli) of one or more modalities (e.g., visual, somatosensory), and asking observers to respond to these stimuli (such as by pressing one of two buttons in a judgement task, see section 3.6). The resultant data are then interpreted to understand the processes underlying perception. A basic example of a psychophysical experiment is the following: Hunter, Godde and Olk (2017) presented gabor patches with varying contrast

luminances, and then plotted the average detection times for each contrast luminance. By fitting a function to this data, an estimate of the relationship between contrast luminance and detection time was obtained. Similar methods are commonly applied throughout the study of perception, using a variety of paradigms. The three most important paradigms for this thesis are simple reaction time (RT) tasks, 2-alternative-forced-choice (2AFC) and adaptive procedures, all of which are discussed in detail in the following sections.

### 3.4. Simple reaction time tasks

The effect of stimulus properties on reaction times can provide insight into the neural processes underlying stimulus processing (Sternberg, 2004). As stated previously, simple RTs reflect both neural processing latencies (Hülsdünker et al., 2019; Schomaker, 2009) and criteria placement (Cao et al., 2007; Watson, 1986), the latter of which is plastic, and can vary based on task demands (Watson, 1986). Further variation in reaction times has been shown to occur due to the observer's past experience with similar tasks (Wong, Goldsmith, Forrence, Haith, & Krakauer, 2017). Some of this inter-individual variability can be accounted for using a repeated measures design, which is recommended for RT research (Sternberg, 2004), and is typically used throughout the experiments in this thesis. By varying certain properties of a stimulus, it is possible to investigate *relative* reaction times, isolating (at least partially) the effect of these properties on relative processing latencies (Sternberg, 2004).

The reaction time experiments contained within the current thesis all contain consistent instructions. Participants were asked to respond as quickly as possible to the detection of a stimulus, be it auditory, visual, or audio-visual. The median was used as a measure of central tendency, since RTs are typically positively skewed and medians are less susceptible to outliers (Whelan, 2008). Practice trials were always included before the reaction time tasks (despite these being relatively simple tasks), since their inclusion reduces variability in RTs in participants with little experience with the task (Sternberg, 2004).

Each of the trials in the RT tasks were presented in random order, and with a randomly varying inter-stimulus interval (ISI). It is important to consider the ISI in experimental designs, since reaction times are dependent on the level of jitter included in stimulus presentations (Wodka, Simmonds, Mahone, & Mostofsky, 2009), and including a jitter has been shown to reduce response time variability (Lee et al., 2015). Furthermore, the inclusion of a randomly varying ISI is important for the EEG experiment described in chapter 6, since variability in stimulus onset time is vital in minimising brain activity associated with anticipation of stimulus onset (Clementz, Barber, & Dzau, 2002). The specific properties of the stimuli in each experiment is discussed in more detail in the upcoming experimental chapters.

### 3.5. Adaptive procedures

Adaptive procedures are often used to determine perceptual thresholds (for a review see Treutwein, 1995), and typically involve the repeated presentation of a stimulus, the level of which usually varies based on a participant's preceding response (or responses). The level of this stimulus varies in 'steps' from an initial value, and therefore one of the most common adaptive procedures is named the 'staircase' procedure. Here, the level of the stimulus typically decreases, either following a specific number of positive responses (typically the detection of a stimulus or a correct response, depending on the experimental design), and increases following a set number of negative responses, which are usually an incorrect detection or an incorrect response. However, the number of positive and negative responses required to shift the stimulus level may be asymmetric (Leek, 2001). The staircase procedure typically aims to determine a specified detection/correct response rate (e.g., 50%), for each observer (Leek, 2001).

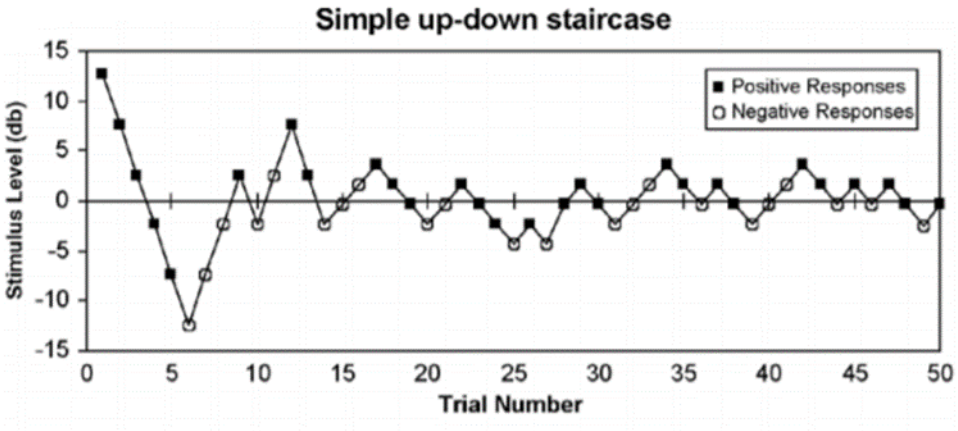


*Figure 3.4 – Typical responses to a 'one-up, one-down' staircase procedure (i.e., the stimulus intensity increases or decreases based on a single positive or negative response), which aimed to determine an observer's perceptual threshold. Figure adapted from Leek (2001).*

An alternative to staircase procedures are maximum-likelihood adaptive procedures, one example of which is the QUEST procedure (Watson & Pelli, 1983). Here, bayesian estimation is used to set the stimulus level of the forthcoming trial, based on an underlying psychometric function that is adjusted following every response (the initial level is based on an estimate of the final threshold value). Consequently, the stimulus level of every trial is set to the value which is currently predicted to be the most likely estimate of the final threshold. This procedure is more efficient than classic adaptive procedures (Watson & Pelli, 1983) and typically avoids presenting trials far from the absolute threshold (Treutwein, 1995). The QUEST procedure was therefore used to estimate the absolute auditory and visual thresholds for observers using the equipment in figure 3.1, in order to determine appropriate visual and auditory intensities to be used in chapters 5 – 7. This was implemented using the QUEST MATLAB toolbox. Two interleaved staircases were presented, in that any single trial could be sampled from the first or second staircase. This is important to ensure that participants cannot anticipate which stimulus will be upcoming

for any particular trial, and consequently adjusting their responses based on this (Cornsweet, 1962).

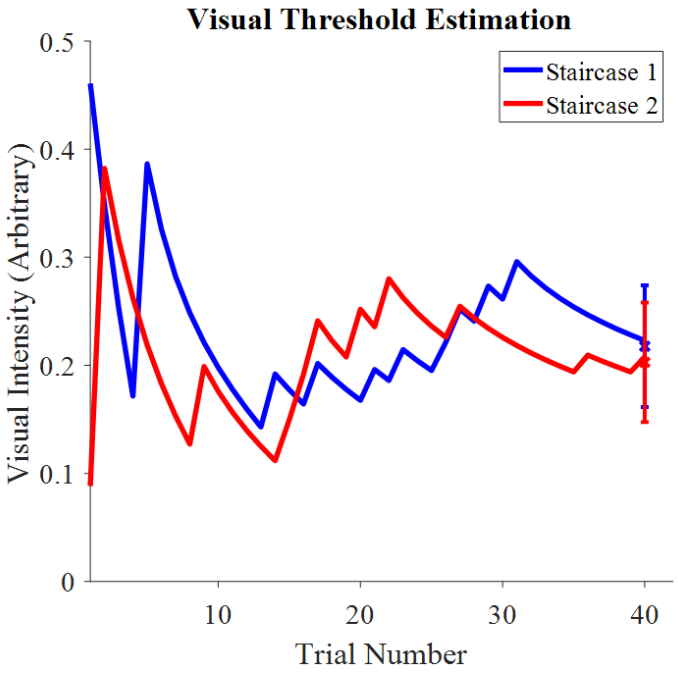


*Figure 3.5 – The estimated absolute visual threshold (+/- SD) of two interleaved staircases of a typical participant (participant three in the preliminary experiment of chapter 5). Each line represents a different staircase, each of which consisted of 40 trials. Staircase 1 started above the final estimated threshold, whereas staircase 2 started below.*

### 3.6. Variants of the SJ and TOJ tasks

The SJ and TOJ tasks were used many times throughout the experiments in the presented thesis. The two tasks are frequently used to assess the relative perceived onsets of audio-visual stimuli (see section 2.2). By manipulating aspects of the experimental design, it is possible to investigate the effects of a variety of exogenous or endogenous factors, such as stimulus parameters or attention, on relative timing perception. It is possible to use identical stimuli across the two tasks, which allows for robust experimental designs, whereby the singular variation between the tasks is the instructions provided to participants.

While adaptive procedures (see section 3.3.1) are typically used to estimate absolute thresholds, they can also be utilised in variants of the SJ or TOJ tasks, where the SOA is varied based on an observer's previous response(s) (García-Pérez, 2014; Wen, Opoku-Baah, Park, & Blake, 2020). Here, the SOA is varied until the procedure converges onto the SOA where simultaneity is most likely to be reported in the SJ task, or the 'light first' and 'sound first' responses are equally likely in the TOJ task, resulting in an estimated PSS. There are some benefits of this, since there is large variability between observer performance on SJ and TOJ tasks (De Niear et al., 2016; Love et al., 2018), and it allows the SOAs presented to observers to be tailored to each individual's ability (Yarrow, 2018).

An alternative to the adaptive procedure variants of the SJ and TOJ tasks is the method of dual presentation, where trials are presented in pairs, and observers must judge, for example, which of the two trials were simultaneous. However, this method has been criticised, since presenting a physically simultaneous stimulus pairing on each trial may lead to observer's adjusting their own *subjective* PSS towards physical simultaneity (Yarrow, 2018).

The most commonly used variant utilises the two-alternative forced choice (2AFC) procedure, where stimuli are presented using the method of constant stimuli, and observers must respond using one of two possible responses after each trial (e.g., simultaneous or non-simultaneous in the SJ, or light first or sound first in the TOJ). With regards to the current thesis, the main benefit of the method of constant stimuli used for the 2AFC variant is that it can be readily applied the RT task, whereas the same does not apply to the method of dual presentation, or to adaptive procedures, where at least two response options are required. Additionally, this method allows for greater flexibility in experimental designs than other methods. For example, it would be difficult to investigate the effect of preceding leading modality on PSS using an adaptive procedure, since

observers would likely be presented with SOAs that are asymmetric around physical simultaneity, which could then shift the PSS by variable amounts in either direction on the following trial through rapid recalibration (see section 2.11.1).

There is a notable drawback that applies to 2AFC, dual presentation and adaptive procedure variants, as PSSs estimated using both the SJ and TOJ tasks have been shown to be impacted by response biases (Keane et al., 2020), including the choice-repetition bias, where observers tend to repeat the response of the immediately preceding trial on the current trial, and the choice-alternation bias, where participants avoid repeating this response (Keane et al., 2020). This could be avoided by asking observers to manually adjust the offset of repetitively presented audio-visual trials, until they are perceived as occurring simultaneously (see Roufs, 1963). However, the stimuli in this variant, again, cannot be readily applied to a simple RT task, nor the TOJ task. Therefore, despite the potential influence of response biases in the 2AFC variant, this was used for the SJ and TOJ tasks presented in the current thesis.

### 3.6.1. The SJ and TOJ tasks used in the current thesis

The trials in the SJ and TOJ tasks in the experiments of the current thesis consisted of audio-visual "flash-beep" stimuli presented across a range of SOAs, that occur in equal amounts in each block and across each observer. It was important to ensure that each SOA was presented an equal number of times to minimise response biases caused by uneven ratios of SOAs (Zerr et al., 2019).

In the SJ task, participants were asked to press one button whenever they perceived the light and sound to have occurred at the exact same time, and a separate button for when they were perceived to have occurred separately. The relative frequency of 'simultaneous' responses was calculated at each SOA, and the resultant data points (see figure 3.6, left) were used for the fitting procedure (see section 3.7). For the TOJ task, participants were

asked to press one button when they perceived the light first and the other when they perceived the sound first, from which the relative frequency of 'light first' responses was calculated at each SOA (figure 3.6, right). For both tasks participants were asked to respond to every trial, and to do this as quickly and accurately as possible.

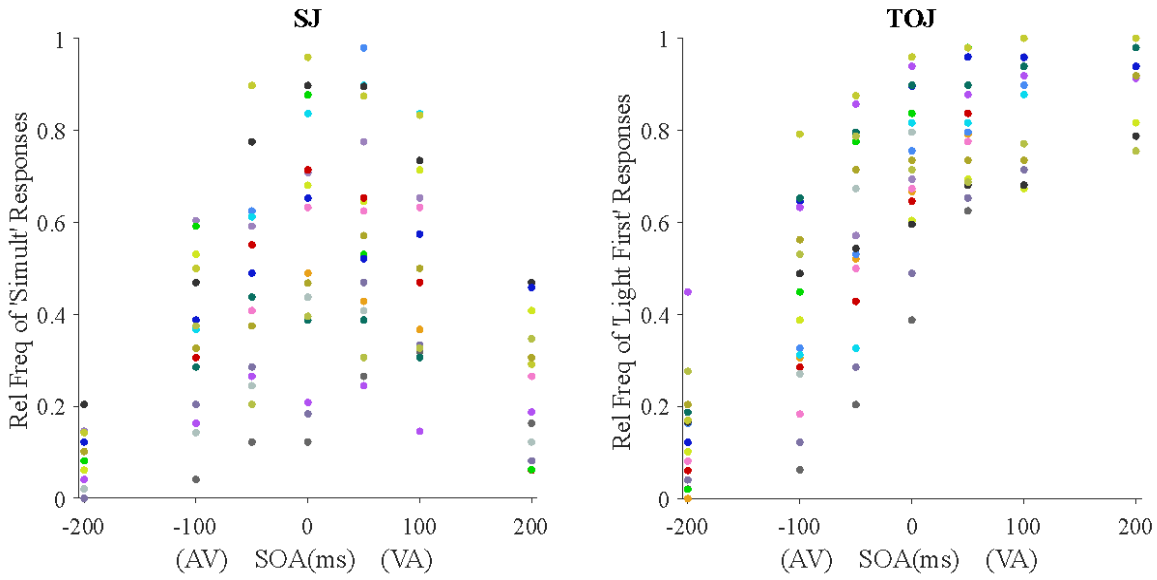


*Figure 3.6 – Example data taken from the experiment presented in chapter 5 (N = 18). Data points represent those collected at the highest visual intensity, in the SJ (left) and TOJ (right) tasks. Each colour represents a different observer.*

Initially (i.e., the experiment presented in chapter 4), the participant's responses were assigned to either the left button ('simultaneous'/'light first') or right button ('non-simultaneous'/'sound first'). There was, however, some confusion from one individual during a practice block of the experiment presented in chapter 4, where they incorrectly pressed the left button as a 'light first' response, despite currently being instructed to complete the SJ task. Therefore, from chapter 5 onwards, a rectangular button box was used which could be rotated between tasks, with the aim of producing a clear distinction for observers in how they should respond for the two tasks. Hereafter, participants were given the button box (see section 3.2) either vertically in the SJ task or horizontally in the

TOJ task. They were instructed to press the top button to report the stimuli as 'simultaneous', the bottom button for 'non-simultaneous' the left button for 'light first' and right button for 'sound first'.

While one of the main aims of this thesis was to investigate the differences between the two tasks, it was also important to utilise both tasks where possible to see if the resultant data from the two tasks converge. Since both tasks are impacted by differing response biases (Linares & Holcombe, 2014; Schneider & Bavelier, 2003), it was important to investigate whether any effect of stimulus properties (e.g., intensity) on PSS was consistent across the two tasks, in order to reduce the likelihood that any reported effect is driven by a task-specific bias.

### 3.6.2. Fitting the SJ data

Past experiments using SJ and TOJ tasks have fitted the data using atheoretical methods, whereby the psychometric function that best reflected the shape of the data was chosen, rather than fitting the data using a method that reflects the researchers' assumptions on the underlying perceptual/psychological processing. Such a method is common in psychophysics (e.g., Leone & McCourt, 2015), and is utilised in chapters 4-6 of the current thesis (details on these fitting procedures are provided in the individual chapters). However, Yarrow, Jahn, Durant and Arnold (2011) developed a fitting procedure that directly represented their theory on how simultaneity and temporal order judgements are made, which was used to fit the SJ data in chapter 7.

Yarrow et al.'s fitting procedure is based on the hypothesis that SJs and TOJs depend both on physical central arrival times, alongside decisional criteria against which the relative arrival times are compared. In other words, it suggests that the PSSs obtained from the tasks do not purely reflect differences in the processing speed of two stimuli. For example, a PSS of -20ms does not necessarily indicate that the visual stimulus was processed 20ms

faster than the auditory stimulus. Instead, this reflects the difference in the processing speed of the auditory and visual stimuli (each with a Gaussian distributed level of sensory noise that effects their respective processing speed) alongside the placement of an individual's criterion (or criteria) along a 'subjective timeline of SOAs'. For TOJs, the perceived SOA (driven by the difference in central arrival time) is compared to an individual criterion in order to judge which stimulus occurred first. Therefore, if an individual's criterion is placed at 20ms (visual-leading), a visual stimulus with a central arrival latency that is 10ms earlier than an auditory stimulus will still be reported as 'sound-leading'. Responses in the simultaneity judgement task are argued to be impacted by the same differential delay (again, driven by differences in processing latency, alongside Gaussian distributed sensory noise). However (importantly for the experiment presented in chapter 7), the relative arrival time is compared against the placement of two decisional criteria.

The importance of decisional criteria in perception has been supported using numerous paradigms (e.g., Aberg & Herzog, 2012; Bang & Rahnev, 2017; Herzog & Fahle, 1999), and research using the SJ task has shown that criterion placement can impact an individual's temporal binding window size (Yarrow & Roseboom, 2017). It has been argued that task manipulations, such as repeatedly presenting stimuli with a consistent visual- or audio-leading SOA, to 'recalibrate' an individual's PSS (i.e., to shift an individual's PSS so that it is closer to the temporal offset that has been repetitively presented) may simply represent temporal shifts in decisional criteria, as opposed to changes in relative perceptual latency (Yarrow et al., 2011).

In chapter 7, the SJ data were fitted using the difference of two cumulative Gaussians (figure 3.7). This results in a 4-parameter fit (*AVμ, AVσ, VAμ, VAσ*), where the means of each curve reflect the mean position of the decisional criteria and the standard deviations reflect the variance in criteria placement, summed with the amount of sensory noise (note

that the individual contributions of the two cannot be disassociated). The size of the temporal binding window has been shown to be asymmetric around physical simultaneity and to vary based on leading modality (Cecere et al., 2016), and therefore it is important to fit the SD of each curve as a separate parameter to allow for this asymmetry.

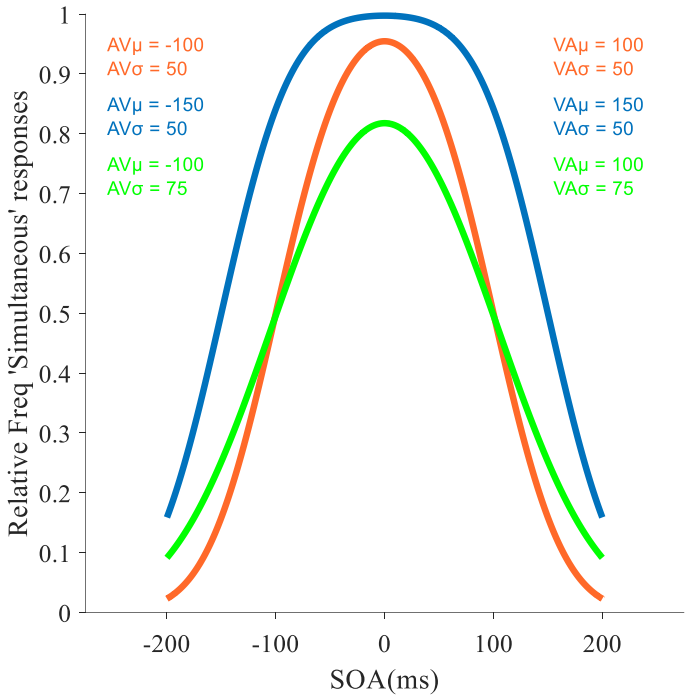


*Figure 3.7 – Simulated curves for the SJ task, produced by varying the four parameters of the fits. To visualise the effect of each parameter, these are displayed for each fit, in text which matches the colour of the respective curve.*

The difference between the mean of each cumulative Gaussian was used as an estimate of the TBW. This method of estimating the TBW is preferred as it normalises the width of the TBW to the amplitude of the curve. An alternative is to calculate the difference between SOAs that correspond to a fixed relative frequency along the y-axis, such as 0.5 (e.g., Butera et al., 2018), however, this method was avoided as it can be affected by the amplitude of the SJ curve, which is heavily influenced by response biases (Keane et al., 2020). For example, an individual who responds with a 'simultaneous' button press whenever they have a lapse in concentration (regardless of physical SOA) would have a higher amplitude,

and therefore a wider estimated TBW than someone who responds with 'non-simultaneous' button presses.

### 3.7. Multisensory Correlation Detector model

Parise and Ernst's (2016) multisensory correlation detector (MCD) model, is based upon the Hassenstein-Reichardt detector model of motion perception (Hassenstein & Reichardt, 1956), where two symmetrical subunits compare inputs from two adjacent receptive fields, after a delay is applied to one of these inputs. The subsequent difference in the signals of the subunits is used to detect the direction of motion. Likewise, The MCD model suggests that the perception of simultaneity and temporal order both rely on two 'detectors' which compare the outputs of two subunits (S1 and S2), the signals in which are driven by the difference in central arrival time of two stimuli. The 'correlation detector' calculates the temporal overlap of the two signals, whereas the 'lag detector' calculates the difference.

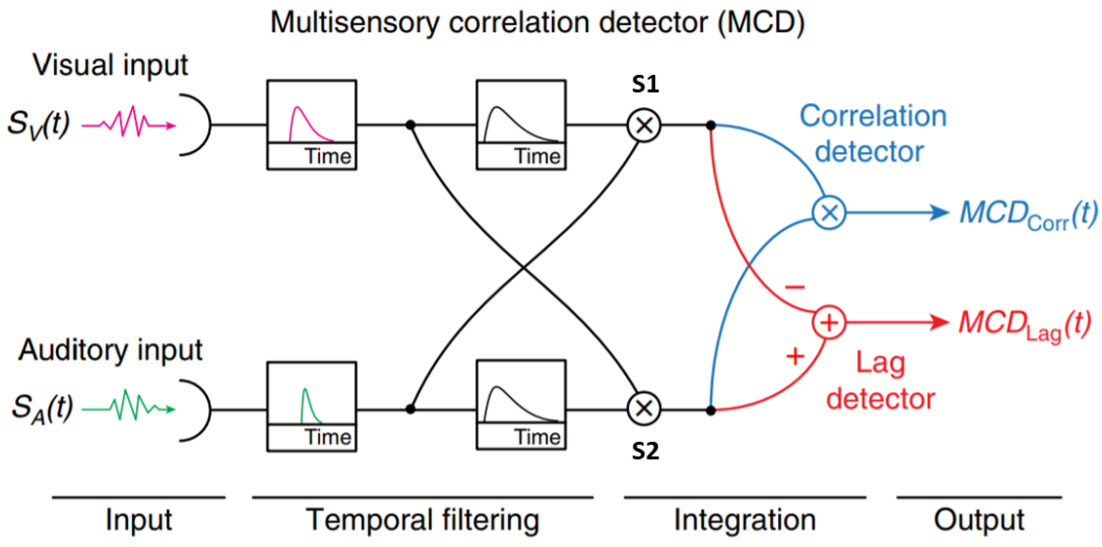


*Figure 3.8 - Schematic figure of the MCD model, adapted from Parise and Ernst (2016).*

In the model, both visual and auditory stimuli pass through modality specific low-pass filters (see figure 3.9): a 87.30ms visual filter ($f_V$) and a 68.40ms auditory filter ($f_A$):

$$f_{mod}(t) = t\ exp(-t/\tau_{mod})$$

whereby $\tau_{mod}$ represents the modality-specific filter time constant and $t$ is time. After one of the signals passes through a second temporal filter: a 785.90ms multisensory filter ($f_{AV}$), the signals are multiplied in S1, S2, which represent two 'mirror symmetric subunits':

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

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

The outputs of the two subunits (see figures 3.10 and 3.11) are then multiplied to produce '$MCD_{CORR}$' whereas the difference of the two subunits produces '$MCD_{LAG}$':

$$MCD_{CORR}(t) = S1(t) . S2(t)$$

$$MCD_{LAG}(t) = S1(t) - S2(t)$$

Note that the calculation of $MCD_{LAG}$ differs from Parise and Ernst (2016) to account for the difference in the shape of the psychometric function fitted to the two data sets (Parise and Ernst fitted a function to the relative frequency of 'sound first' responses, as opposed to 'light first'). Finally, the output of the detectors is averaged over a 5 second window, and the resultant mean is divided by the sum of the signals after they have passed through the respective modality-specific filters in order to normalise the signals (thereby nullifying the effect of input signal 'intensity' on the correlation and lag detector, see section 3.7.1):

$$\overline{MCD_{CORR}} = \frac{mean[MCD_{CORR}(t)]}{\sum[S_A(t) * f_A(t)] + \sum[S_V(t) * f_V(t)]}$$

$$\overline{MCD_{LAG}} = \frac{mean[MCD_{LAG}(t)]}{\sum[S_A(t) * f_A(t)] + \sum[S_V(t) * f_V(t)]}$$

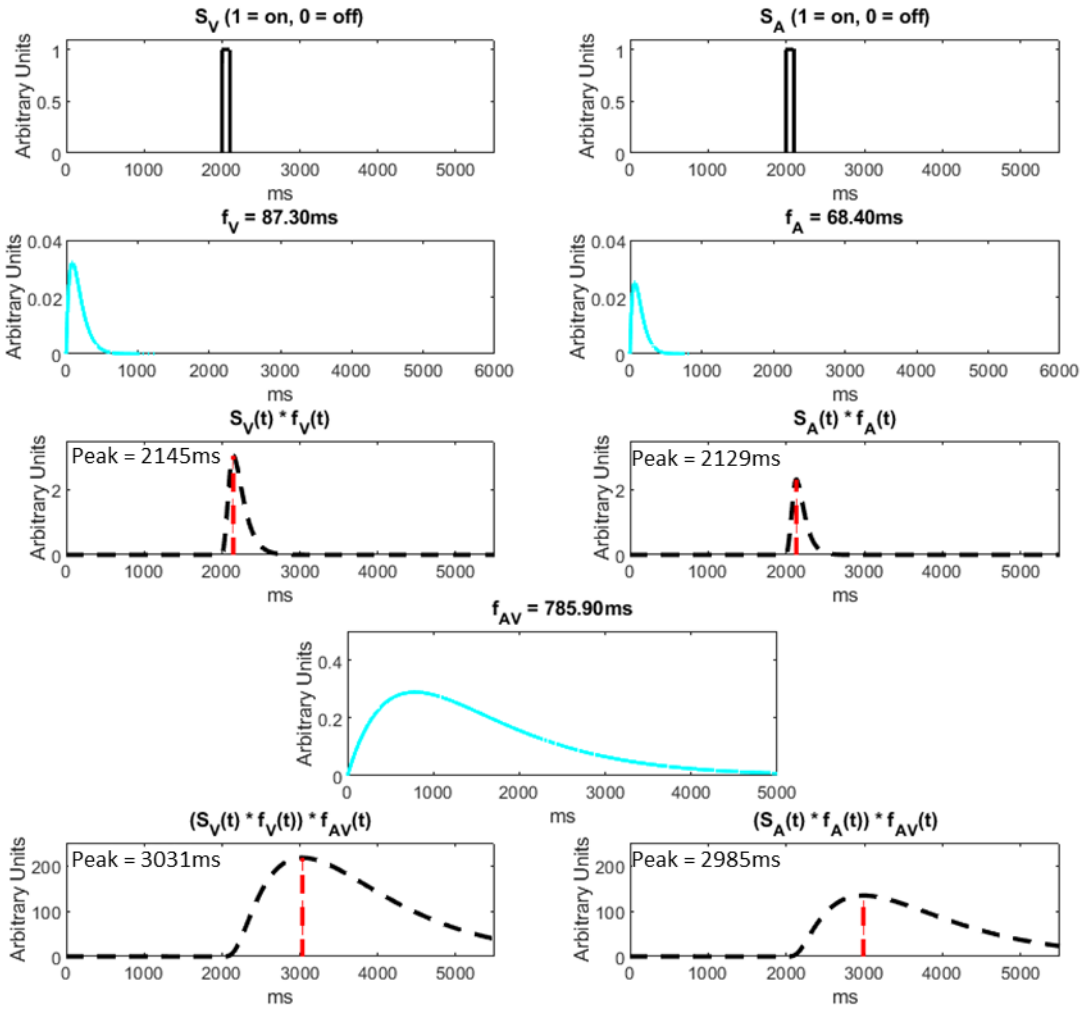


*Figure 3.9 - The processing of simultaneously presented visual (left column) and auditory (right column) stimuli, produced using simulated data. The first row contains the two stimuli, which are 100ms binary signals presented at 2000ms for clarity. The second row displays the visual and auditory filters. The difference in the time constants of the filters accounts for the longer processing latencies for visual compared to auditory stimuli. The two signals are convolved with their respective filters, producing a visual signal with a later peak than the auditory signal (row 3). The resultant signals are each convolved separately with the multisensory filter (row 4), which produces the signals in row 5. Each subunit receives the signal of one modality that has passed through Mf, and one that has not. The function of the Mf is to delay one of the two signals before it reaches a subunit (S1 or S2), analogous to the delay to a visual signal in a Reichardt detector.*

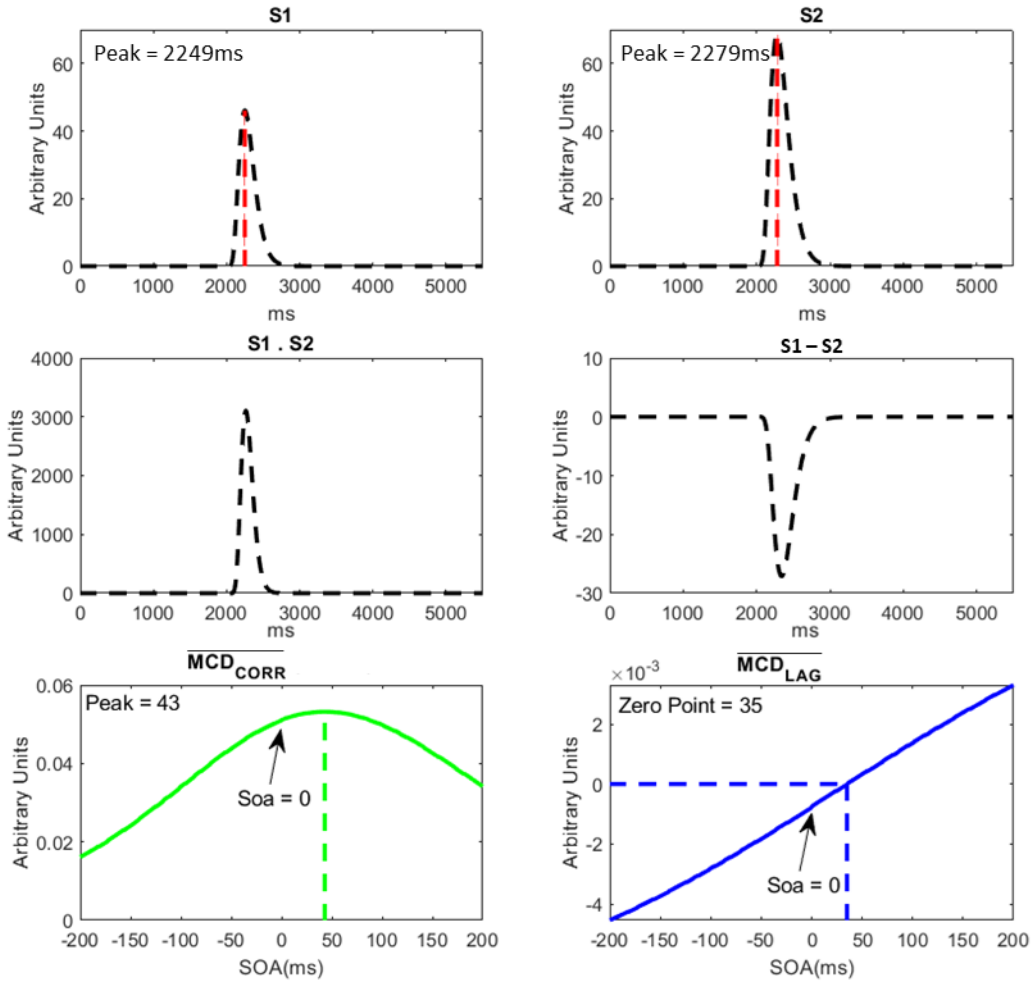


*Figure 3.10 - The first row represents the signals in subunit 1 (left) and subunit 2 (right), produced from simultaneously presented (SOA = 0) audio-visual signals. Subunit 1 represents the product of the visual signal convolved with $f_V$ and $f_{AV}$, and the auditory signal convolved with $f_A$. Subunit 2 represents the product of the auditory signal convolved with $f_A$ and $f_{AV}$, and the visual signal convolved with $f_V$. Row 2 displays how $MCD_{CORR}$ (left) and $MCD_{LAG}$ (right) are calculated for a given SOA: the correlation detector (left panel), which calculates the product of both subunits, and the lag detector (right panel), which takes the difference between the subunits. Row 3 represents the time averaged model outputs over a range of SOAs (with the SOA of the signals presented in the rows above highlighted with an arrow).*

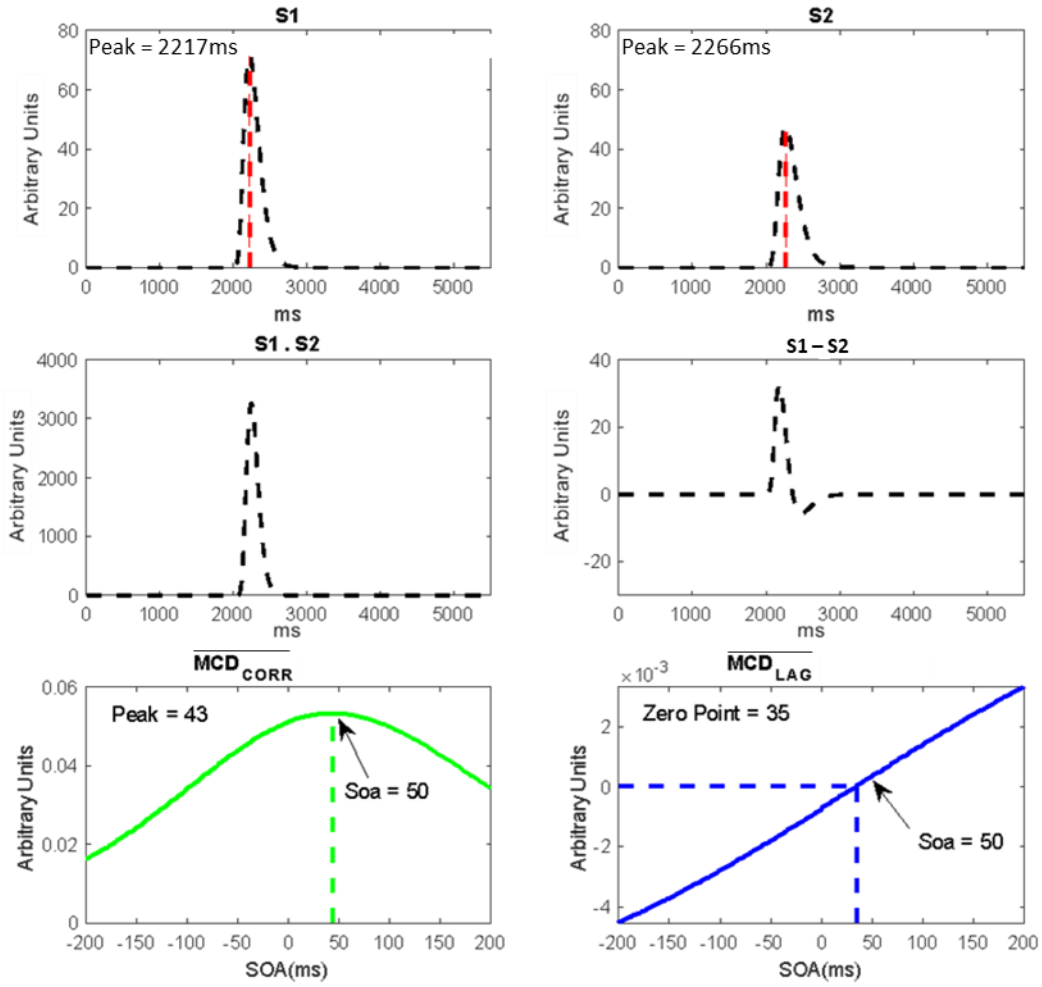


*Figure 3.11 – The processing of a 50ms visual-leading signal, between the subunits and the correlation and lag detectors. Note that the earlier processing of two offset signals is mostly identical to figure 3.9. The temporal offset simply causes a temporal shift in the signals, both before and after they pass through the filters, whereas their size and shape remain constant. As can be seen from row 1, subunit 1 (left) now has a larger signal than subunit 2 (right). This leads to a positive $\overline{MCD_{LAG}}$ output at 50ms, reflected in the time averaged model outputs displayed in row 3. The product of subunits 1 and 2 produces a slightly larger signal, and therefore a larger $\overline{MCD_{CORR}}$ output, compared to an SOA of 0ms (of course, the shape of $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ outputs presented across all SOAs (bottom row) are identical to that seen in figure 3.10).*

$\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ are mapped into to the behavioural data (Y) for the SJ and TOJ tasks separately, via general linear model with a probit link function. Importantly, both model outputs are used as inputs, regardless of whether the SJ or TOJ data is being fitted (see figure 3.12):

$$F(Y) = Y' = \beta_0 + \beta_{CORR} * \overline{MCD_{CORR}} + \beta_{LAG} * \overline{MCD_{LAG}}$$

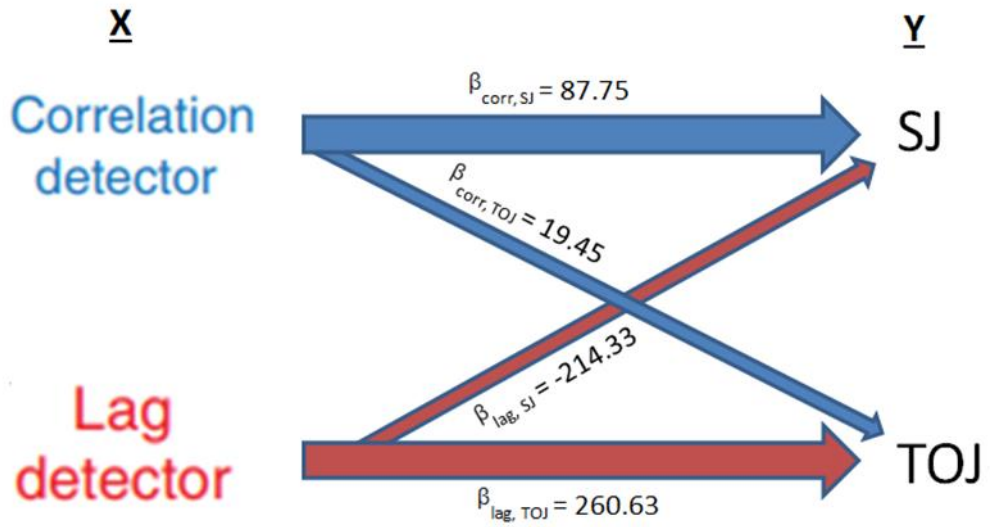


*Figure 3.12 - The weightings (beta values) of the correlation detector and lag detector when the high visual intensity behavioural data taken from chapter 4 were fitted using the MCD model. A general linear model with a probit link function was fitted to the behavioural data (Y) for both tasks (SJ and TOJ), with $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ as inputs (X).*

### 3.7.1 The effect of intensity on PSS – a limitation of the MCD model

A limitation of the MCD model is that it cannot account for the known effect of stimulus intensity on processing latency (Carrillo-de-la-Pena et al., 1999) and on the PSS (Roufs, 1963). Figure 3.13 displays the time averaged model outputs of the MCD model when the coded 'intensity' of the visual input (defined by the amplitude of the input signal) is manipulated without normalisation (see section 3.7). The model does not predict a shift in PSS. Instead, the amplitude of the correlation detector increases with increasing intensity.

The same increase in amplitude of a curve fitted to actual SJ data would reflect an increase in the reporting of simultaneity across all SOAs, which is indicative of a response bias, rather than a PSS shift (Linares & Holcombe, 2014). Furthermore, the steeper slope seen in the lag detector output would reflect an increase in perceptual sensitivity in the TOJ task (van Eijk et al., 2008), rather than a PSS shift. Logically, if the MCD model were to predict the effect of visual intensity on the PSS (Leone & McCourt, 2015), one might expect the arrival time of the visual signal at each of the two subunits to be dependent on the intensity of the visual stimulus. Higher visual intensities, with relatively shorter processing latencies, would therefore be expected to cause a shift in the peak of the $\overline{MCD_{CORR}}$ output and the zero point of the $\overline{MCD_{LAG}}$ output towards the auditory leading SOAs.

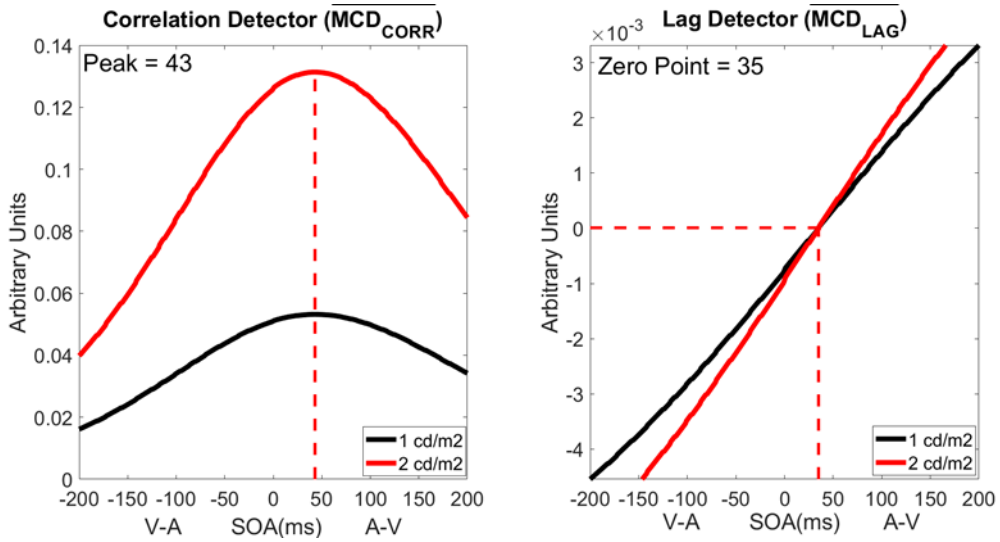


*Figure 3.13 – The effect of varying the inputted stimulus intensity on the $\overline{MCD_{CORR}}$ (left) and $\overline{MCD_{LAG}}$ (right) values.*

### 3.8. Electroencephalography

Electroencephalography (EEG) is a neuroimaging method used to measure the electrical activity of neurons, providing insight into a wide array of neurocognitive processes in the brain (e.g., Louis & Frey, 2016). This electrical activity (measured in microvolts) is generated from extra-cellular electrical charge that is produced by the flow of ions through neuron membranes (Teplan, 2002), which leads to distortions of the electrical field that are

detected by electrodes that are placed on the scalp (Beres, 2017). The signal must be amplified, as it is attenuated by the various layers it has to pass through before reaching these electrodes, including the skull, which is an especially poor conductor (Lanfer et al., 2012). EEG has many applications and has been used previously in a wide range of clinical and experimental settings, ranging from the study of perception (Hramov et al., 2017; Tayeb et al., 2020), to the diagnosis of epilepsy (Angus-Leppan, 2008; Smith, 2005).

There are several benefits of using EEG over other neuroimaging methods, the most notable being its extremely high temporal resolution, which is in the order of milliseconds (e.g., Burle et al., 2015), meaning the resultant data generally reflects the time course of neurocognitive processes more closely than other neuroimaging methods, such as fMRI and positron emission tomography (PET) (Teplan, 2002). Furthermore, the EEG trace is a direct measure of neural activity, in comparison to fMRI which measures changes blood oxygen levels that are assumed to be associated with neural activity but can also be influenced by other bodily processes (Chen & Li, 2012). Other benefits of EEG include the fact that this is non-invasive, meaning that there is little-to-no risk for participants, its cheap in comparison to most similar methods, and it is also relatively portable, meaning that EEG is a highly flexible method to study the brain. There are some drawbacks of using EEG however, most notably of which is the poor spatial resolution relative to other neuroimaging methods, such as fMRI and PET (Teplan, 2002), as signals often have to pass through several layers of neural tissue from their source, making the precise location of this source difficult to identify (Väisänen, 2008). Therefore, EEG is typically used to assess activity in the cerebral cortex (Teplan, 2002), largely because sources from subcortical structures are much more difficult to detect than from superficial regions (Samuelsson, Khan, Sundaram, Peled, & Hämäläinen, 2019).

### 3.8.1. Event-related potentials

EEG experiments often involve the study of event-related potentials (ERPs). These are electrical changes in electroencephalographic recordings that are time-locked with sensory or cognitive events. ERP components were traditionally believed to be driven by synchronous activity from large groups of neurons (Peterson, Schroeder, & Arezzo, 1995), however, more recent research has highlighted the important role of oscillatory phase reset on the ERP (Fell et al., 2004; Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007). Further details on the role of neural oscillations in ERPs are beyond the scope of this thesis, but for a detailed review see Klimesch et al. (2007).

ERP components reflect peaks in the evoked potential that are commonly named based on the polarity (Positive, 'P' or Negative, 'N'), and the order in which they occur. For example, the first positive component of the visual evoked potential is termed the 'P1', though it should be noted that the timing of the peak of the component is used interchangeably with its order in EEG-based literature, for example, the P1 may be referred to as the 'P100'. There are ERPs that are associated with the perception of various stimuli, alongside the performing of different tasks. Trials involving these stimuli and response combinations are usually repeated many times. By time-locking the EEG data to stimulus onsets (or any other time point in relation to the stimulus, or to a participant's response) in ERP designs, the EEG data of many trials are then typically averaged within specified time windows (referred to as epochs), reducing some of the noise inherent with the use of EEG (Luck, 2014). The resultant averaged data is the ERP, which is typically compared across conditions to investigate the effect of these conditions on the neural response.

The fact that EEG is subject to a large amount of noise guided the experimental design of the EEG experiment contained in this thesis (chapter 6), in that the number of SOAs at which trials were presented at was reduced (relative to the equivalent behavioural experiment discussed in the chapter), to allow for more trials to be presented at these

SOAs, therefore reducing the signal to noise ratio when the ERPs were averaged (Luck, 2014).

### 3.8.2. Typical ERP components and their sources

Due to its high temporal resolution, electroencephalography has been used to make valuable contributions to the understanding of the time courses of visual and auditory processing streams. The earliest components of the visual evoked potential (VEP) comprise the C1, a negative potential that occurs as early as 55ms after stimulus onset and is generated in the primary visual cortex (Clark, Fan, & Hillyard, 1994), the P1 component (which onsets around 80ms), originating from the dorsal extrastriate cortex, and the N1 component (with an onset between 110-120ms), originating from the ventral extrastriate cortex (Di Russo, Martínez, Sereno, Pitzalis, & Hillyard, 2002).

The time course of early auditory processing is evident in the P1-N1-P2 complex (Schröder et al., 2014), comprising the P1 (with an onset of around 50ms), arising from the primary auditory cortex, the N1 (onsetting around 100ms), which is generated by both the primary and secondary (A2) auditory projection areas, and the P2, which is generated solely in A2 (with an onset around 200ms) (Winkler, Denham, & Escera, 2013).

Multisensory evoked potentials typically reflect the additive neural activity of their unisensory components, which are then modulated by the introduction of a second stimulus (Besle et al., 2004). Importantly, it is this modulation (i.e. the deviation from summation) that is reflective of multisensory integration (Besle et al., 2004). For example, both the auditory N1 and visual N1 are modulated by the introduction of a second stimulus, with more negative amplitudes observed for bimodal trials, in comparison to the sum of its unimodal components (Brandwein et al., 2011). Likewise, the auditory P2 is reportedly reduced in amplitude for bimodal stimuli (Klucharev et al., 2003), whereas the amplitude of

the visual P1 is increased for bimodal stimuli (Giard & Peronnet, 1999), again, in comparison to the sum of their unisensory components.

### 3.9. Chapter Summary

The current chapter described various methods commonly used to study audio-visual perception, and justified the design choices used in the experiments presented throughout this thesis. It also summarised the equipment which was purpose built by the author, the design of which was driven by the experimental procedures described in the upcoming chapters.

Two tasks commonly used to study relative timing judgements are the simultaneity and temporal order judgement tasks. The processes underlying these tasks will be investigated throughout this thesis, using both the psychophysical and electrophysiological methods described in this section. While both of these tasks on the surface appear very similar, and differ only in the instructions given to participants, they often produce uncorrelated findings, provoking several researchers to produce processing models to explain these differences. The MCD model described above is a particularly interesting model, however it cannot (at least in its current form) explain the intensity dependency of the PSS. This effect of intensity is investigated in the following chapter (chapter 4), and a modified version of this MCD model is presented in chapter 5.

# 4.Pilot experiment

## 4.1. Chapter introduction

The relative central arrival latencies of audio-visual stimuli are determined by the physical offset in their arrival latency at the appropriate sensory organs, and subsequently by their physiological processing latencies. Processing latencies are intensity dependent (see section 1.6), and it is therefore unsurprising that visual intensity impacts the physical offset required between an audio and visual stimulus for them to be perceived as simultaneous (Roufs, 1963). Since the integration of audio-visual stimuli is dependent on the temporal overlap of the comprising unimodal signals (Meredith & Stein, 1986), it would perhaps be expected that varying the relative intensity of two signals would also impact the physical offset between the signals required for them to be integrated. However, Leone and McCourt (2015) showed that an intensity manipulation which led to a PSS shift, estimated using the SJ and TOJ tasks, did not impact the SOA where facilitative multisensory integration was observed. It was concluded that intensity dependent differences in processing latencies are compensated for, via coincident detectors which take into account the relationship between relative arrival latencies and stimulus intensity. Importantly, it was argued that this process was specific to the RT task (Leone & McCourt, 2015). Since Leone and McCourt (2015) used a small sample in their experiments (n = 5 and n = 6), the current chapter aimed to replicate their findings using a much larger group of observers. Furthermore, this chapter also aimed to replicate the uncorrelated PSS estimates reported between the SJ and TOJ tasks (e.g., Van Eijk, et al., 2008), with the ultimate aim of expanding on this finding in later chapters.

### 4.1.1. The effect of temporal alignment on multisensory enhancement

External events commonly result in the production of multimodal cues, and correctly determining which external stimuli occurred from the same source is vital to gain an accurate view of the world around us. One of the cues used to decide whether an audio-visual stimulus was produced from the same physical event is its temporal coincidence (Burr, Silva, Cicchini, Banks, & Morrone, 2009). Unisensory signals must propagate through independent channels in order to reach multisensory regions, such as the superior colliculus, where they converge onto individual 'multisensory' neurons (Meredith, 2002). The temporal alignment of these signals has an important role in multisensory processing, as it leads to greater firing rates in these multisensory neurons (Meredith et al., 1987), an effect termed 'multisensory enhancement' (Meredith & Stein, 1986). Violations of the race model inequality (Miller, 1982) in simple RT data (consisting of both bimodal and unimodal trials) can be used as a behavioural indicator of multisensory enhancement (Gondan & Heckel, 2008; Gondan & Minakata, 2016). Here, it can be assumed that bimodal signals have been integrated when the redundant signals effect is too large to be attributable to the fastest of the two unimodal signals (measured as the sum of their distribution functions, see section 4.2.7).

### 4.1.2. The effect of relative intensity on multisensory enhancement

Both visual (Carrillo-de-la-Pena et al., 1999; Lines et al., 1984) and auditory (Jaskowski et al., 1994) processing latencies are dependent on stimulus intensity, reflected in reduced reaction times to higher intensity visual and auditory stimuli (Jaśkowski, 1992; Jaskowski et al., 1995; Jaskowski et al., 1994; Ulrich et al., 1998; van der Molen & Keuss, 1979). While multisensory enhancement is most likely to occur for temporally aligned stimuli, it may be expected that by varying the relative intensity of an audio-visual stimulus, it would be possible to shift the SOA at which the race model inequality is violated *away* from physical

simultaneity. In other words, if the intensity of a visual stimulus is increased whilst the auditory stimulus is held constant, the resulting effect on the relative arrival latencies at multisensory neurons would mean that a physically simultaneous stimulus pair would no longer be integrated. Yet Leone and McCourt (2013) found this to be untrue. They presented audio-visual stimuli across a range of stimulus onset asynchronies (SOAs) and audio-visual intensities, and reported significant violations of the race model for physically simultaneous stimuli *regardless* of relative stimulus intensity.

### 4.1.3. A proposed mechanism to compensate for intensity-dependent processing latencies in facilitative multisensory enhancement

Leone and McCourt (2013; 2015) argued that differences in the arrival latencies of visual and auditory signals at multisensory neurons, that are caused by differences in the relative intensities of the two signals, are corrected for. They proposed that through sensory experience, the synaptic connections of coincidence detectors are tuned to audio-visual stimuli of a specific intensity. Physically simultaneous audio and visual signals of specific intensities will only be integrated at a coincidence detector tuned to these intensities when they arrive with a specific latency differential (i.e. a differential that reflects the relative intensity dependent SC arrival times of signals produced by physically simultaneous stimuli) (Leone & McCourt, 2013). They proposed this process likely occurs at the multisensory neurons in the superior colliculus, who's distribution (Xu, Sun, Zhou, Zhang, & Yu, 2014) and capability to integrate multisensory stimuli (Xu, Yu, Stanford, Rowland, & Stein, 2015) have both been shown to be dependent on early life sensory experiences.

#### 4.1.3.1. The dissociation between 'action' and 'perception' subsystems

While simple RTs are employed to measure the effect of intensity on perceptual latencies, the simultaneity judgement (SJ) and temporal order judgement (TOJ) tasks, in theory, provide a measure of perceived *relative* latencies (see chapter 2). In contrast to the

evidence of multisensory integration for physically simultaneous stimulus, regardless of intensity manipulations (Leone & McCourt, 2013; 2015), there appears to be a consistent effect of intensity on *perceived* relative latencies, as estimated using the SJ and TOJ tasks. For example, as visual intensity is decreased, the PSS shifts further towards the visual leading SOAs in the SJ (Leone & McCourt, 2015; Menendez & Lit, 1983; Roufs, 1963) and TOJ (Jaśkowski, 1992; Leone & McCourt, 2015) tasks.

Leone and McCourt (2015) proposed that the simple RT task is mediated by an 'action' subsystem which compensates for intensity dependent differences in processing speed, whereas the SJ and TOJ tasks are mediated by a 'perception' subsystem, in which no such compensation takes place. The discrepancy between the two posited subsystems is supported by Harrar, Harris and Spence (2016), who presented repetitive temporally offset trials to recalibrate the observer's perception of simultaneity. Despite the pre-test adaptation phase shifting the observers' PSSs, the race model was violated exclusively for simultaneous stimuli in RT trials which followed the same adaptation phase. Thus, there is reasonable evidence for a discrepancy in the 'action' and 'perception' tasks.

#### 4.1.4. Task differences in estimates of perceived simultaneity

Research has shown that the SJ and TOJ tasks produce conflicting results, despite these tasks theoretically measuring similar processes. For example, the PSS of the TOJ task has been shown to be typically positioned in the audio-leading SOAs (Zampini et al., 2003), whereas the PSS of the SJ task typically lies in the visual-leading SOAs (Zampini, Guest, Shore, & Spence, 2005). Furthermore, it has been reported that there is a lack of correlation between the PSSs estimated using the SJ and TOJ tasks (Basharat et al., 2018; Binder, 2015; Linares & Holcombe, 2014; Love et al., 2013; Van Eijk et al., 2008; Vatakis et al., 2008). These differences have led to the proposal that the SJ and TOJ tasks are dependent on separate underlying mechanisms (Zampini et al., 2003; Love et al., 2013).

### 4.1.5. Rationale and aims

To date there is reasonable evidence for a dissociation between the 'action' (i.e. RT) and 'perception' (i.e. SJ and TOJ) tasks. However, Leone and McCourt's (2015) experiments had very small sample sizes (n = 5 for the RT task, n = 6 for the SJ and TOJ tasks) and made no comparison between the SJ and TOJ tasks, despite previous evidence that they can produce uncorrelated estimates. The current chapter therefore aimed to replicate the inconsistent effects of intensity on the 'action' and 'perception' tasks (Leone & McCourt, 2015), with a considerably larger sample size. Furthermore, it aimed to replicate the previously reported lack of correlation between the PSSs estimated using the SJ and TOJ tasks.

In a repeated measures design, 38 observers completed the audio-visual SJ, TOJ and RT tasks, with a combination of bimodal and unimodal stimuli. A constant auditory intensity, and two visual intensities were presented to observers in separate blocks.

### 4.1.6. Predictions

It was predicted that, in line with past research, facilitative multisensory integration in the RT task would occur at physical simultaneity for both visual intensities. Secondly, it was predicted that the PSS recorded from the SJ and TOJ tasks would be significantly higher (i.e. shifted towards the visual-leading offsets) in the low visual intensity condition compared to the high visual intensity condition, reflecting the relatively slower processing speeds for less intense stimuli. Finally, it was predicted that the PSSs, recorded at both the low and high visual intensities, would be uncorrelated between the SJ and TOJ tasks.

## 4.2. Method

### 4.2.1. Participants

38 participants (20 female) aged 20-69 (mean = 39.42, SD = 16.52) were obtained through opportunity sampling, via advertisements on the University of Liverpool's website.

Participants reported normal or corrected-to-normal vision, and normal hearing. 4 participants were excluded from all final analyses, due to their estimated PSS lying outside the presented range of SOAs in either the SJ or the TOJ task at either the high or low visual intensity conditions.

### 4.2.2. Design

A repeated measures design was used. Each participant completed the SJ, TOJ and RT task at both a high and low visual intensity condition, in a pseudo-random order, on two separate days. Participants would never complete the same task twice on one day (e.g., the simultaneity judgement task would be completed under the low visual intensity condition on the first day, and then under the high visual intensity condition on the second). The order of the SJ and TOJ tasks were counterbalanced, so that these were always the first and last tasks for each participant in a given session. The second task would always be the simple RT task to give the participant a break from the more cognitively demanding judgement tasks.

### 4.2.3. Apparatus

The study was conducted in a soundproof booth (IAC, Winchester, UK). Identical audio-visual flash-beep stimuli were used for all 3 tasks. The stimuli were presented using the Arduino Genuino Micro, connected to a Windows computer running the Arduino software (version 1.6.6). The visual stimuli were presented using a simple red LED, whereas the audio stimuli were presented using a piezo sounder. The LED was located 1cm above the piezo sounder, both of which were located exactly 1m away from the participant's sitting position. The auditory stimulus was produced with a fixed frequency of 1000Hz at 68db, with a flat tone amplitude envelope and a 100ms duration. The visual stimulus was presented for 100ms at 1.1 cd/$m^2$ in the low intensity condition and 366 cd/$m^2$ in the high

intensity condition. Participants responded to the stimuli using a two-button computer mouse that was connected directly to the Arduino.

### 4.2.4. Stimuli

Each participant completed 3 blocks at both the high and low visual intensities in each of the 3 tasks. Trials were presented across 9 bimodal SOAs (-200, -150, -100, -50, 0, 50, 100, 150 and 200ms, where negative values indicate audio-leading trails), alongside unimodal visual and auditory trials. It should be noted that the unimodal trials were included in all tasks for consistency, but were only analysed for the RT task. Within each block, each of the bimodal SOAs and the two unimodal stimuli were presented 12 times each in pseudo-random order. As a result, there were 36 trials per SOA across all blocks, for both the high and low intensity conditions, totalling 792 trials per participant, per task.

The gap between each stimulus presentation was randomised between 1000 and 3000ms to minimise anticipation and fast guesses, which is especially important for the reaction time task (Gondan & Minakata, 2016). Participants were given 2000ms to respond to each trial before the next stimulus was presented. For the SJ and TOJ tasks, responses earlier than 100ms after stimulus onset were removed, however this did not apply to the RT task to maximise the statistical power of the RT analysis (Gondan & Minakata, 2016).

### 4.2.5. Procedure

Observers were seated in the soundproof booth and given instructions based on the upcoming task. For the SJ task, observers were asked to press the left mouse button if they perceived the audio-visual stimuli to have occurred simultaneously, and the right mouse button if they were non-simultaneous. ‘Non-simultaneous’ responses should also be given for unimodal stimuli. In the TOJ task they were asked to press the left mouse button if the stimulus was perceived to be visual-leading, and the right button if it was perceived to be

auditory-leading. Observers were instructed to also press the left button for unimodal visual trials, and the right for unimodal auditory trials. For the reaction time task, observers were instructed to press the left mouse button as quickly as possible whenever they detected a unimodal or bimodal stimulus. Before each of the three tasks the instructions were repeated through the open door of the soundproof booth. Participants were then given a single practice block to get accustomed to the buttons, followed by a chance to ask questions. They were then given 10 minutes of dark adaptation before the experiment began.

#### 4.2.6. Data fitting and analysis – SJ and TOJ data.

The relative frequency of 'simultaneous' and 'light first' responses were calculated in the SJ task and TOJ task, respectively. This was done at each SOA for each observer. To obtain the best-fitting parameters for the SJ and TOJ data, a nonlinear least-squares method was used (Function 'fit'; MATLAB2018). A four parameter Gaussian function was fitted to the SJ data:

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

where $a$, is the offset along the y-axis, $b$ is the amplitude of the symmetric Gaussian, $c$ is the midpoint and $d$ is the width. The parameter $c$ was used as the estimate of the PSS.

For the TOJ data, a four parameter sigmoidal curve (Boltzman) was fitted:

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

where $a$ and $b$ reflect the upper and lower asymptotes of the function, respectively, $c$ reflects the location of the curve, and $d$ reflects the width. The SOA corresponding to a 50% response rate of 'light first', was used as the estimate of the PSS.

To investigate the effect of intensity on PSS, repeated measures *t-tests* were used to compare the PSSs estimated for the two visual intensity conditions. This was done

separately for the SJ and TOJ tasks. The Pearson correlation coefficient was used to assess the correlation between PSSs of the SJ and TOJ tasks at the 2 visual intensities.

### 4.2.7. Reaction time analysis and race model inequality test.

To ensure the intensity manipulation impacted unimodal visual RTs, Wilcoxon signed ranks tests were used to compare the median RTs for the low and high intensity unimodal visual stimuli.

Violations of the race model inequality were used as evidence of facilitative multisensory integration in the simple RT task. To test the race model, individual reaction times were divided into 30ms time bins (Harrison et al., 2010). For the bimodal stimuli, this was done separately for each SOA and visual intensity. Likewise, this was done separately for the unimodal auditory trials, and the two intensities of the unimodal visual trials. The cumulative distribution frequency at each time bin, between 0 and 300ms, was calculated (as a speeded response to bimodal stimuli is most likely to occur in these early time bins). The cumulative frequency of the bimodal SOAs were compared with the summed cumulative frequency of the two unimodal conditions, using the following equation taken from Gondan and Minakata (2016):

$$P\left(T_{AV} \leq t\right) \leq \left[P\left(T_A \leq t\right) + P\left(T_V \leq t\right)\right]$$

where $P$ represents the cumulative probability density function of reaction time ($T$) to either the bimodal ($AV$), auditory ($A$) or visual ($V$) stimulus at a specific time bin ($t$).

Because reaction times have a positively skewed distribution, a Wilcoxon signed ranks test was used to statistically assess whether the race model was violated within any given time bin.

## 4.3. Results

### 4.3.1. Effect of intensity on simple reaction times

Observers completed a simple RT task, in response to bimodal audio-visual trials across varying SOAs and two visual intensities, and to unimodal audio and visual trials at the same intensities. The averages across each observer's median RTs are presented in figure 4.1. The leftmost and rightmost bars represent the unimodal audio and visual RTs, respectively. For the unimodal visual trials, RTs were significantly faster for high intensity stimuli (239.90ms) than for the low intensity stimuli (280.60ms) ($Z$ = -4.59, $p$ < .001), confirming an effect of intensity on RTs.

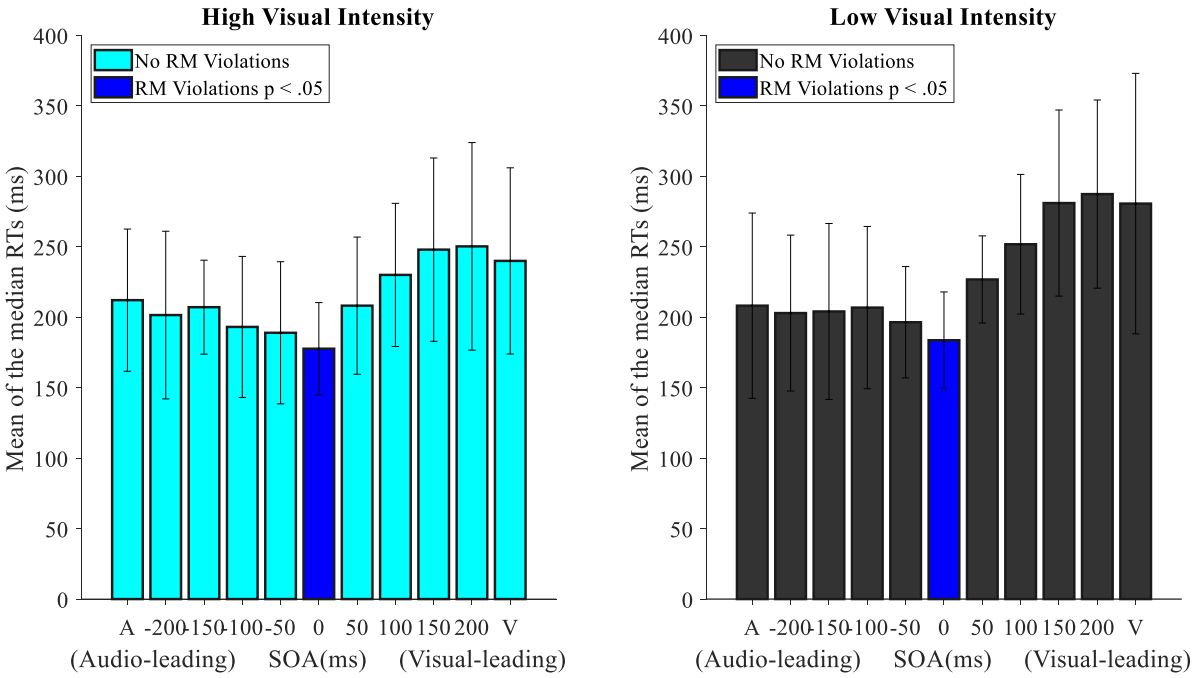


*Figure 4.1 – The average across each observer's median bimodal and unimodal reaction times in the high intensity (left) and low intensity (right) conditions. Error bars represent the interquartile range. The SOAs where significant violations of the race model were found are represented by dark blue bars. The manipulation of visual intensity appeared to primarily affect the visual-leading reaction times, which is unsurprising assuming RTs are dictated by*

*the modality with the earliest central arrival time in these temporally offset trials (Miller, 1982).*

To investigate the effect of intensity and temporal overlap on facilitative multisensory integration (as evident through RTs), the race model inequality was tested at each SOA for high and low intensity trials separately. The analysis revealed significant violations of the race model inequality exclusively at an SOA of 0ms for both the high and low intensity conditions, suggesting that facilitative multisensory integration occurred only when the two stimuli occurred simultaneously. These significant violations were at time bin 5 (120-150ms) ($Z$ = 2.67, $p$ = .008) and time bin 6 (150-180ms) ($Z$ = 2.56, $p$ = .010) in the high intensity condition, and time bin 6 in the low intensity condition ($Z$ = -3.27, $p$ = .001). No significant violations were found at any other SOA or time bin (p > .05).

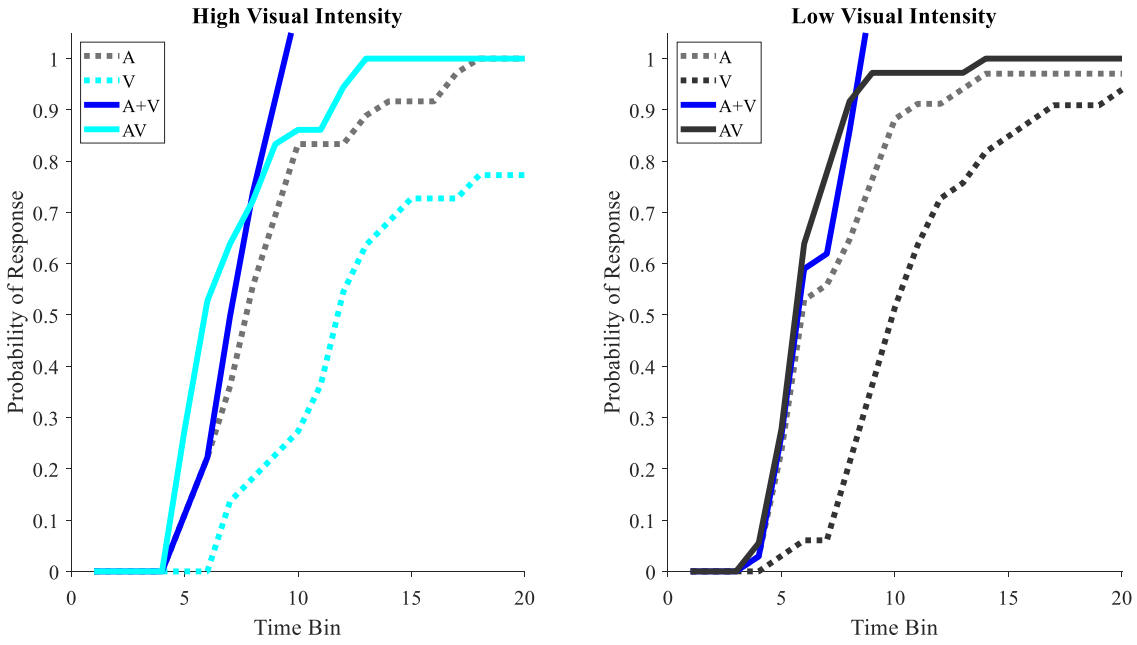


*Figure 4.2 – Cumulative distribution functions for a typical participant (participant 4) in the high intensity (left) and low intensity (right) condition at SOA = 0ms.*

### 4.3.2. Effect of visual intensity on Perceived Simultaneity

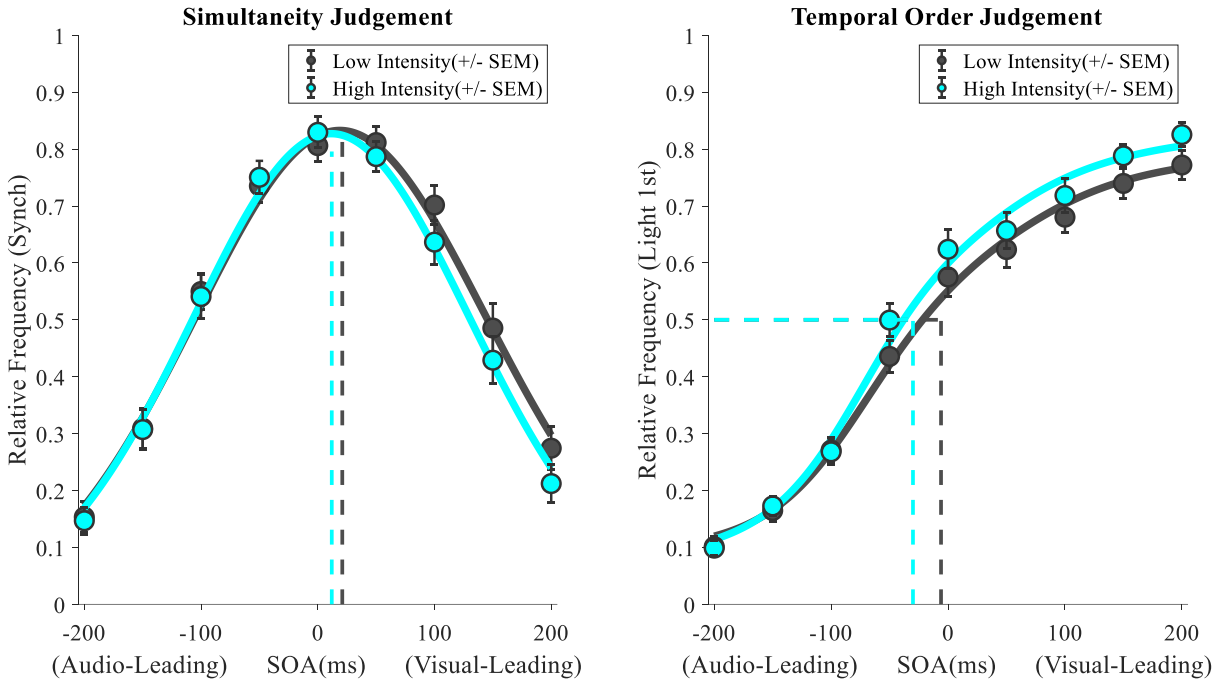


*Figure 4.3 – Mean fits for the SJ (left) and TOJ (right) tasks for the high (cyan) and low (grey) intensity conditions. Circular markers and error bars reflect mean relative frequencies and the standard error of the mean, respectively. Dashed lines reflect the mean PSSs.*

The mean fits across observers, in the audio-visual SJ and TOJ tasks at the two visual intensities are plotted in figure 4.3. For the SJ task, there was a significant difference between the PSSs estimated in the high ($M$ = 10.64ms, $SD$ = 26.93) and low ($M$ = 19.93ms, $SD$ = 23.30) visual intensity conditions ($t$ (33) = 3.18, $p$ = .003). Likewise, for the TOJ task there was also a significant difference between the PSSs in the high ($M$ = -30.35 ms, $SD$ = 50.54) and low intensity ($M$ = -6.21ms, $SD$ = 61.46) conditions ($t$ (33) = 2.20, $p$ = .035). For both tasks, the PSS was higher in the low intensity condition, meaning that as the intensity of the visual stimulus is reduced, the visual flash must occur earlier, relative to the auditory 'beep', in order for simultaneity to be reported.

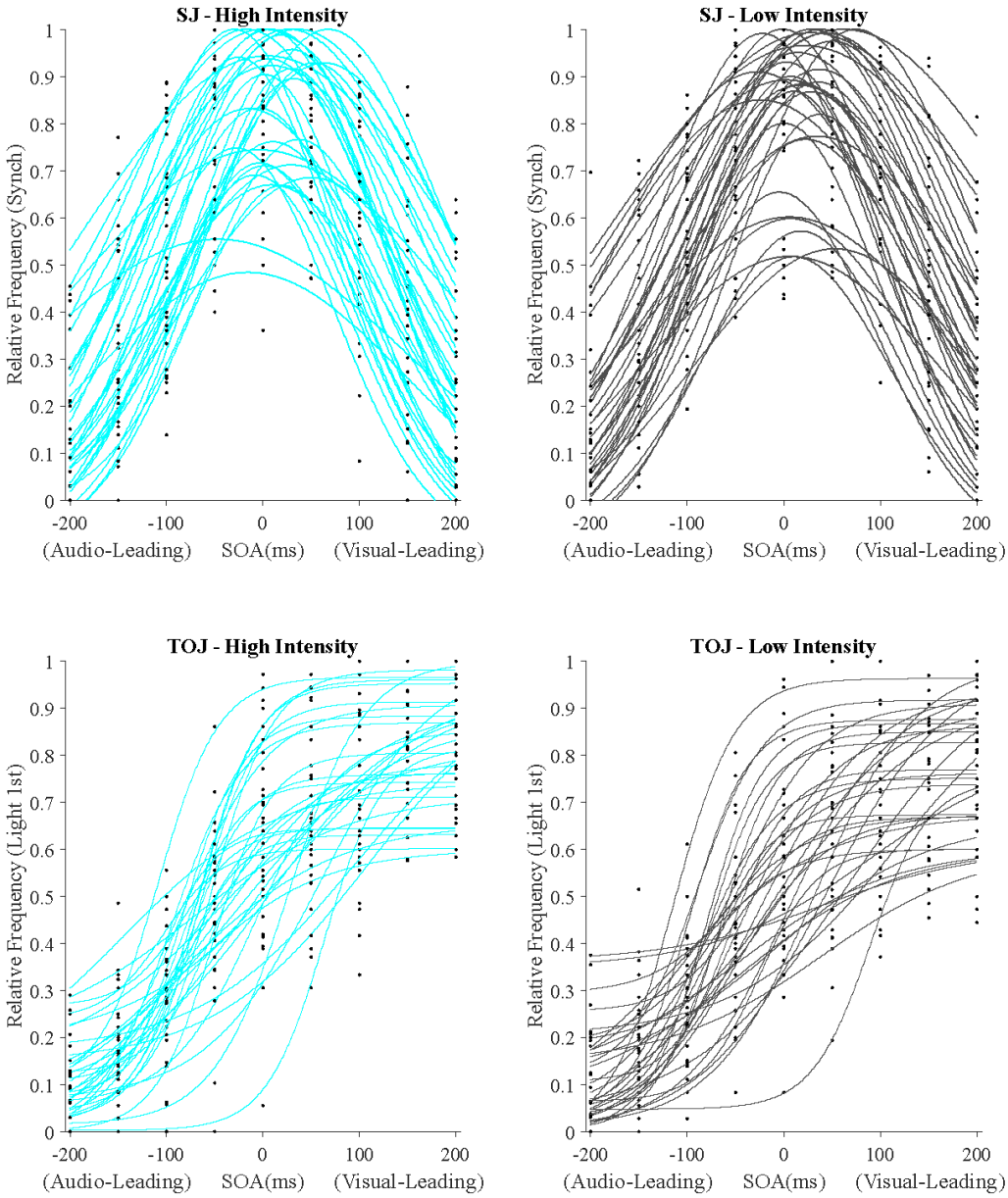


*Figure 4.4 – Data points (black circles) and fits (grey or cyan lines) for individual observers in the SJ (top row) and TOJ (bottom row) tasks, at the high (left column) and low (right column) visual intensities.*

The individual PSSs for the SJ and TOJ task are plotted in figure 4.5. A relatively larger spread in the PSSs estimated using the TOJ task is evident, perhaps reflecting the higher subjective difficulty of the task, in comparison to the SJ task (Love, 2013). The Pearson correlation coefficient revealed that there was no significant correlation between the PSSs of the SJ and TOJ tasks in both the high intensity condition ($r$ (32) = -.17, $p$ = .328) and low intensity condition ($r$ (32) = -.24, $p$ = .173).

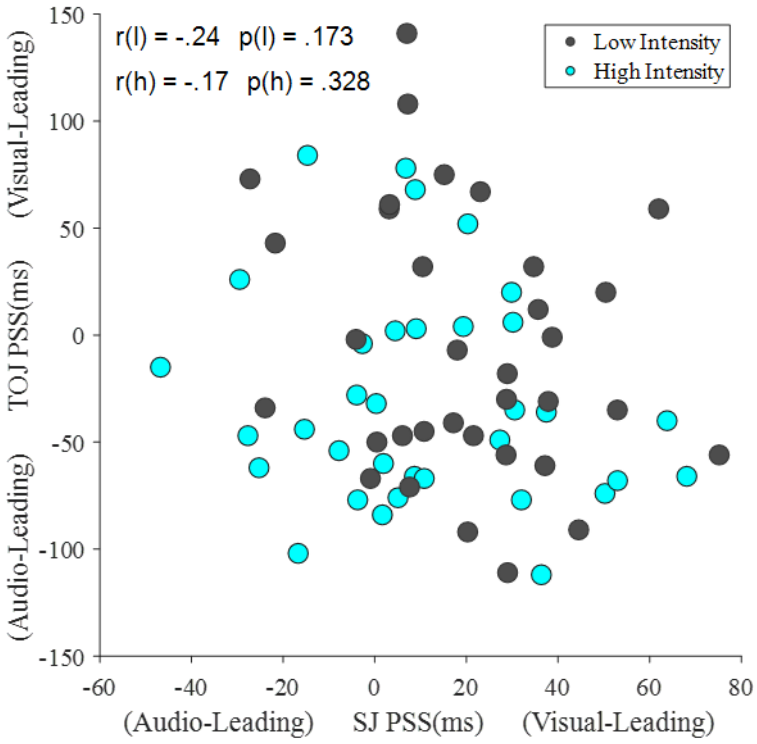


*Figure 4.5 – Scatterplot displaying the PSSs estimated using the SJ and TOJ tasks. Values reflect the Pearson correlation coefficient at the low (l) and high (h) visual intensities.*

### 4.4. Discussion

Observers completed the audio-visual SJ, TOJ and simple RT tasks at two visual intensity levels, to investigate whether the previously reported inconsistent effect of stimulus intensity on perceived simultaneity and facilitative multisensory integration (Leone &

McCourt, 2015), could be replicated. The data from the SJ and TOJ tasks showed that the PSS shifts towards the visual-leading SOAs in the low intensity condition, relative to the high intensity condition. In contrast, the RT data showed that violations of the race model occurred exclusively at an SOA of 0ms regardless of visual intensity. The results are therefore consistent with those reported by Leone and McCourt (2015) and the predictions of the current chapter. A second aim of the current experiments was to replicate the previously reported uncorrelated PSSs, estimated using the SJ and TOJ tasks. The analysis revealed no significant correlation, at either the high or low visual intensity conditions, again consistent with previous research and the predictions of the current chapter.

#### 4.4.1. Divergent effects of visual intensity on perceived simultaneity and facilitative multisensory integration

The data from the SJ and TOJ tasks showed that intensity had a significant effect on perceived simultaneity. Compared to the high visual intensity trials, the low intensity stimulus had to come earlier, relative to the auditory stimulus, in order for simultaneity to be perceived. Since the SJ and TOJ tasks provide an estimate of relative perception latencies (Linares & Holcombe, 2014), the current data is consistent with the idea that perceived simultaneity is influenced by intensity dependent processing latencies (Leone & McCourt, 2015; Roufs, 1963). While relative stimulus intensity effected the PSS, the RT analysis revealed that race model violations occurred only when the audio-visual stimuli were presented at physical simultaneity (SOA = 0ms), *regardless* of visual intensity.

The results presented here are in line with Leone and McCourt's (2013; 2015) experiments, who provided tentative mechanism that could explain why facilitative multisensory integration was evident at an SOA of 0ms, regardless of relative stimulus intensity. They argued that through coincident detectors, differences in convergence latencies at multisensory neurons that are caused by variations in stimulus intensity, can be

compensated for (Leone & McCourt, 2015). While the current data cannot confirm the existence of such a mechanism, they do provide some support for the distinction between the effect of intensity on the 'action' (simple RTs) and 'perception' (SJ and TOJ) tasks, with a substantially larger sample size. It should be noted however, that the mean PSS shifts observed in the current data were relatively small (< 50ms). Assuming that the observed PSS shifts are driven by intensity-dependent processing latencies, it is possible that the intensity manipulation used in the current experiments, while large enough to cause a shift in the relatively sensitive PSS estimates, was not large enough to impact the SOA at which the race model was violated (estimated in intervals of 50ms). Nevertheless, Leone and McCourt (2013) used smaller SOA intervals (20ms), yet still reported a consistent violation of the race model inequality for physically simultaneous stimuli. Furthermore, Leone and McCourt (2015) reported a 75ms shift in the PSS estimated in their SJ task, yet still reported significant violations of the race model (measured in intervals of 50ms) exclusively at an SOA of 0ms. Therefore, when taken together with the current findings, there is behavioural evidence for the compensation of intensity dependent processing latencies, specifically during the RT, but not the SJ or TOJ tasks.

#### 4.4.2. Proposed benefit of independent processing mechanisms for 'action' and 'perception' tasks

While the underlying benefit of a theorised distinction between the processing mechanisms underlying 'action' and 'perception' tasks is currently unknown, a mechanism which compensates for relative intensity differences would ensure that the behavioural (Leone & McCourt, 2013; Stevenson et al., 2012) and neural changes (Meredith et al., 1987) associated with simultaneous central arrival latencies occur for *physically* simultaneous stimuli, when rapid reaction times are required (Leone & McCourt, 2015). Such a process

could be vital for reduced RTs to ecologically valid signals, with comparable arrival times at peripheral sensors, that may be vital for survival.

While there is an obvious benefit of coincident detectors which compensate for the relationship between stimulus intensity and relative central arrival times, the reasoning for why this same mechanism would not compensate for the effect of intensity on perceived simultaneity is unclear. For example, the temporal offset between audio-visual signals is vital in determining the identity of perceived spoken syllables (ten Oever, Sack, Wheat, Bien, & van Atteveldt, 2013). Since auditory processing latencies are intensity dependent (Jaskowski et al., 1994), the variations in amplitude which occur naturally in speech (e.g., Arnold & Watson, 2015), will likely impact the latency differential of audio-visual signals at multisensory neurons, perhaps therefore disrupting speech perception (ten Oever et al., 2013). Thus, it is unclear as to why coincident detectors would not also compensate for intensity-dependent processing latencies when making perceptual judgements on relative timing (resulting in a PSS that is unaffected by relative intensity), when these judgements are so important to social processes, such as speech perception.

Further research involving neuroimaging methods with a high level of spatial and temporal resolution, such as a combination of fMRI and EEG (Mullinger & Bowtell, 2010), would be required to assess when and where the processing streams of the 'action' and 'perception' tasks differ, and to rule out the possibility of discrepant decisional criteria (e.g., Yarrow, 2018) causing the inconsistency seen in the data.

#### 4.4.3. Divergent PSS estimates from the SJ and TOJ tasks

Alongside the reported disparity between the action and perception tasks, the current data also highlights the difference between the two 'perception' tasks. In line with past research (Love et al., 2013; van Eijk et al., 2008), the PSSs estimated using the SJ and TOJ tasks were found to be uncorrelated for both the high and low visual intensity stimuli.

Since the SJ and TOJ tasks have been repeatedly used together to study various disorders (e.g., Capa et al., 2014; Regener et al., 2014; Scarpina et al., 2016; Wise & Barnett-Cowan, 2018), the current results emphasise the need to understand the differences in the processing underlying the two. Again, future research using neuroimaging techniques is necessary to discern whether these differences are due to separate underlying processing mechanisms (e.g., Zampini et al., 2005), or due to differences at the decisional stage (e.g., Parise & Ernst, 2016).

### 4.5. **Chapter summary**

To conclude, the experiments reported in the current chapter replicate the findings reported by Leone and McCourt (2015). The results highlighted an effect of intensity on the PSS, yet regardless of visual intensity, facilitative multisensory integration was evident only when the stimuli were presented simultaneously, consistent with the proposal that intensity-dependent latency differences are compensated for when performing simple RTs, but not when making SJs or TOJs. Furthermore, the PSSs estimated using the SJ and TOJ tasks were uncorrelated, consistent with past research. Together, the results further emphasise the noteworthy difference, not only between the so-called 'action' and 'perception' tasks, but also between the two 'perception' (SJ and TOJ) tasks.

As well as the inconsistency between the effect of intensity on the PSS and on facilitative multisensory integration, previous research has also highlighted an inconsistency in the effect of intensity on the PSS and *unimodal* simple RTs (Jaśkowski, 1992; Menendez & Lit, 1983). However, due to the use of only two visual intensities in the current chapter, the intensity dependency of the PSS and simple RTs could not be explored. The following chapter therefore builds upon the current experiment by utilising several visual intensities, and a constant auditory intensity.

The MCD model predicts the uncorrelated PSSs estimated using the SJ and TOJ tasks but cannot account for the effect of intensity on the PSS reported in the current chapter. The following chapter therefore presents an extended MCD model that aims to account for this effect of intensity.

# 5. Visual intensity-dependent response latencies predict perceived audio–visual simultaneity

## 5.1. Chapter introduction

The previous chapter investigated the effect of visual intensity on SJs, TOJs and simple RTs. This revealed an effect of intensity on the PSS, with a shift towards the auditory-leading SOAs for more intense visual stimuli, suggestive of a role of intensity-dependent processing latencies. However, previous research has highlighted an effect of visual intensity on simple RTs that is roughly twice the magnitude of the respective effect on the PSS (Jaśkowski, 1992; Menendez & Lit, 1983), perhaps implicating a role of higher-order processes in the effect of intensity on the PSS, which have been argued to ensure subjective perception aligns more closely with the timing of physical events (Harris, Harrar, Jaekl, & Kopinska, 2010). However, this discrepancy may also be related to attentional effects which reduce the effect of intensity on the PSS (Jaśkowski, 1992) or by intensity-related differences in the average perceived temporal offset in each block of the SJ or TOJ tasks, which can bias responses (Miyazaki et al., 2006), potentially reducing the effect of intensity (see section 2.9). Importantly, these latter possibilities depend on the visual stimulus intensity only being manipulated across blocks. The current chapter therefore utilised the audio-visual SJ, TOJ and simple RT tasks, with multiple visual intensities that were interleaved *within* blocks. This aimed to investigate the shared role of visual intensity-dependent processing latencies in the effect of intensity on all three tasks.

Furthermore, the previous chapter replicated the uncorrelated PSS estimates of the SJ and TOJ tasks. The MCD model parsimoniously accounts for this discrepancy in PSS estimates,

however this model cannot account for the effect of intensity on the PSS. With the aim of assessing the role of early and late processing stages in the observed intensity dependency, an intensity-dependent pre-processing stage was added to this model, and several model variants were compared.

## 5.2. 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.

### 5.3. 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).

### 5.3.1. 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 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] 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 ($AIC_c$).

### 5.3.2. 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.

### 5.3.3. 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 5.1). 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.

## 5.4. A modified Multisensory Correlation Detector

### 5.4.1. 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 5.1): 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 5.1). The time averaged $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ detector outputs are then mapped into behaviourally obtained psychometric data (SJ or TOJ task) via a probit link function:

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

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

where Y denotes the psychometric function obtained in the SJ and TOJ tasks, respectively and F is the probit link function. The coefficients $\beta_{CORR}$ and $\beta_{LAG}$ are estimated from the psychometric functions and are fitted for each data set (TOJ, SJ) separately, but assuming the same detector inputs.

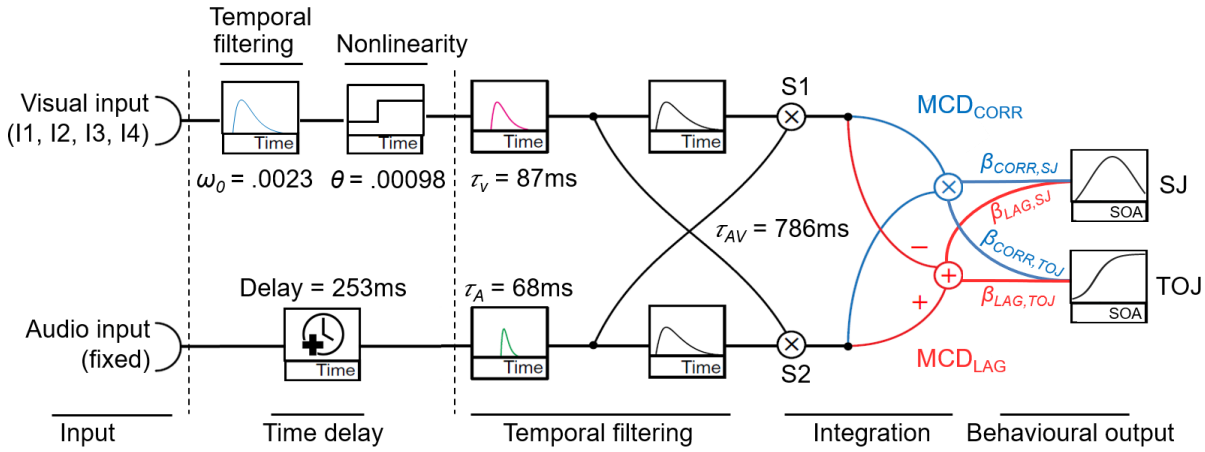


*Figure 5.1. The Multisensory correlation detector (MCD) model (Parise & Ernst, 2016) with the added early visual-intensity dependent component (within dashed lines). The filter time constants ($\tau_V$, $\tau_A$, and $\tau_{AV}$) were taken from the original MCD model and are not optimised for our data set. The parameters of the early component (filter cut-off $\omega_0$ and threshold $\theta$) 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 $\overline{MCD_{CORR}}$ output for the SJ and TOJ data; similarly, $\beta_{LAG,SJ}$ and $\beta_{LAG,TOJ}$ denote the weights of the $\overline{MCD_{LAG}}$ 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 $\beta$ optimised for each intensity; in models 1 and 3 the coefficients $\beta$ are shared parameters (see Table 5.1 and also Table 5.C.1).*

To model the known intensity-dependence of perceived simultaneity, we included an additional early processing stage (Temporal filter and subsequent nonlinearity; Figure 5.1), 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 ($\tau_V$, $\tau_A$, *and* $\tau_{AV}$) and only the coefficients $\beta_{CORR,SJ}$ *and* $\beta_{CORR,TOJ}$ (Figure 5.1) are estimated from our behavioural TOJ and SJ judgements (see 3.1.2: Experiment). The parameters of the intensity-dependent early component ($\omega_0$ *and* $\theta$; Figure 5.1, left side) are estimated from a separate set of reaction time data (see section 5.4.3).

### 5.4.2. Variants of the MMCD model

| 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 | $\omega_0$, $\theta$; $n$, $t_{max}$, $RT_0$ | 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 | - $\omega_0$, $\theta$; $n$, $t_{max}$, $RT_0$ | 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 5.1 –The 4 model variants. See text for details.*

To evaluate the modified MCD model, we fit four different nested variants of the model (Table 5.1). Model 1 does not allow for any intensity dependence of perceived simultaneity and the three coefficients $\beta$ are estimated for all intensities simultaneously. In Model 2 the coefficients $\beta$ 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 $\beta$ are shared

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

### 5.4.3. 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 5.1). 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 5.2 (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 B). 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 $\omega_0$, the integration time $t_{max}$, threshold $\theta$ and the irreducible reaction time $RT_0$. (See table 5.2). The most relevant parameter is the number of filtering stages (n), as seen in Figure 5.2. A filter order of n=6 captures the linear relationship between reaction times and log intensity well. The cut-off $\omega_0$ and the

threshold $\theta$ cannot be estimated independently from each other since the absolute error is similar for different parameter combinations (see Figure 5.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.

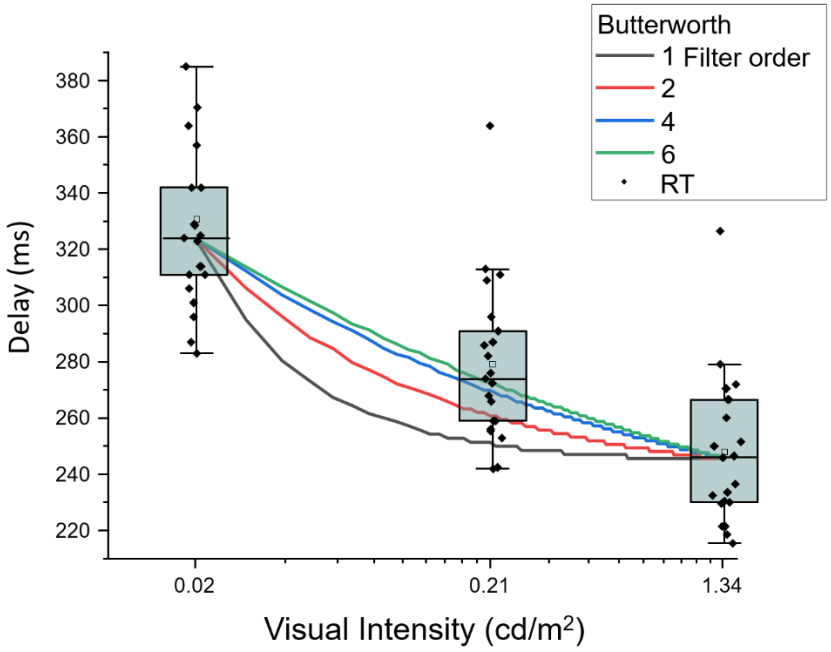


*Figure 5.2: Median RT time data (black diamonds) are used to estimate the parameters of the Butterworth filter. Outputs for filters of order 1,2,4 and 6. The experimental data show individual RTs as well as median and upper/lower quartiles as box plots. The open symbols show the mean RTs. The optimisation used the median RTs (horizontal lines in boxes). The model output is quantised into milliseconds.*

| Butterworth filter | |
|---|---|
| Cut-off $\omega_0$ | 0.0023 ($f_c$= 1.15Hz; $f_s$=1kHz) |
| order $n$ | 6 |
| integrator $t_{max}$ | 100ms |
| threshold $\theta$ | 0.00098 |
| $RT_0$ | 207ms |

*Table 5.2: Filter and thresholding parameters. Note that there is no unique solution for the filter cut-off $\omega_0$ and threshold $\theta$ since they co-vary (see also figure 5.B1).*

This time-delayed visual signal then feeds into the MCD (Figure 5.1). 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.

### 5.4.4. 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 5.3) for different visual intensities. The SOAs corresponding to the absolute peak and absolute minimum of the $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ 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 5.3, 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 5.6.2).

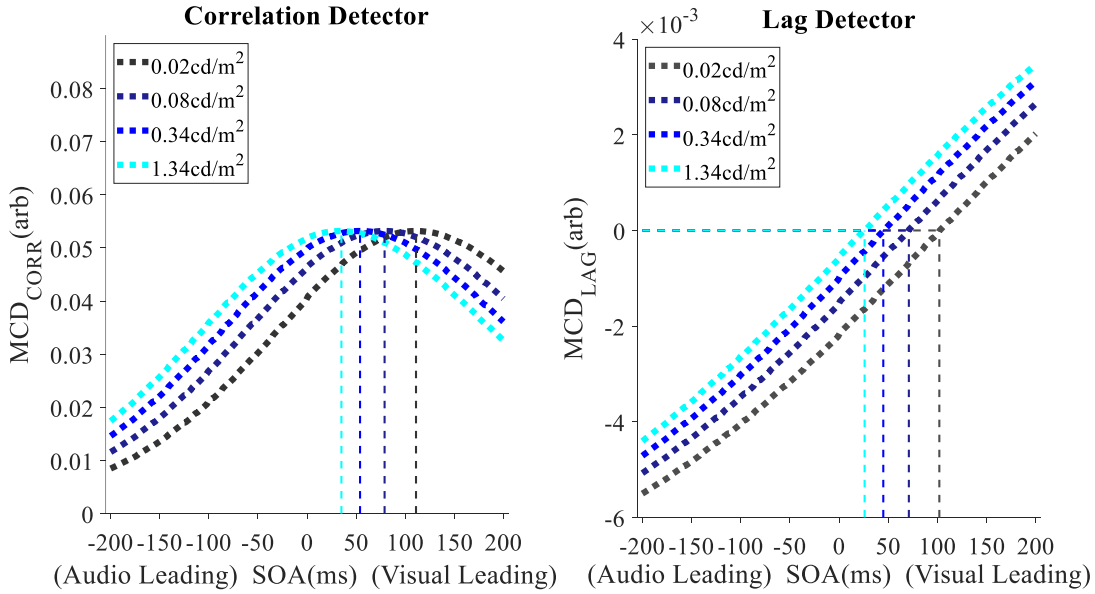


*Figure 5.3. Outputs of the correlation ($\overline{MCD_{CORR}}$ - left) and lag detector ($\overline{MCD_{LAG}}$ - right) for the 4 visual intensities. The SOAs corresponding to the absolute peak and absolute minimum of the $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ detector outputs, respectively, 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 $\overline{MCD_{CORR}}$ peaked at 43ms whereas the SOA corresponding to a $\overline{MCD_{LAG}}$ value of 0 was 35ms.*

### 5.5. 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.

### 5.5.1. Methods

### 5.5.2. 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/$m^2$.

### 5.5.3. 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.

#### 5.5.4. 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.

## 5.6. Results

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

### 5.6.1. Psychometric functions for different intensity levels

The individual fits for each intensity level are shown in Figure 5.4 (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 5.4a). 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 5.4b).

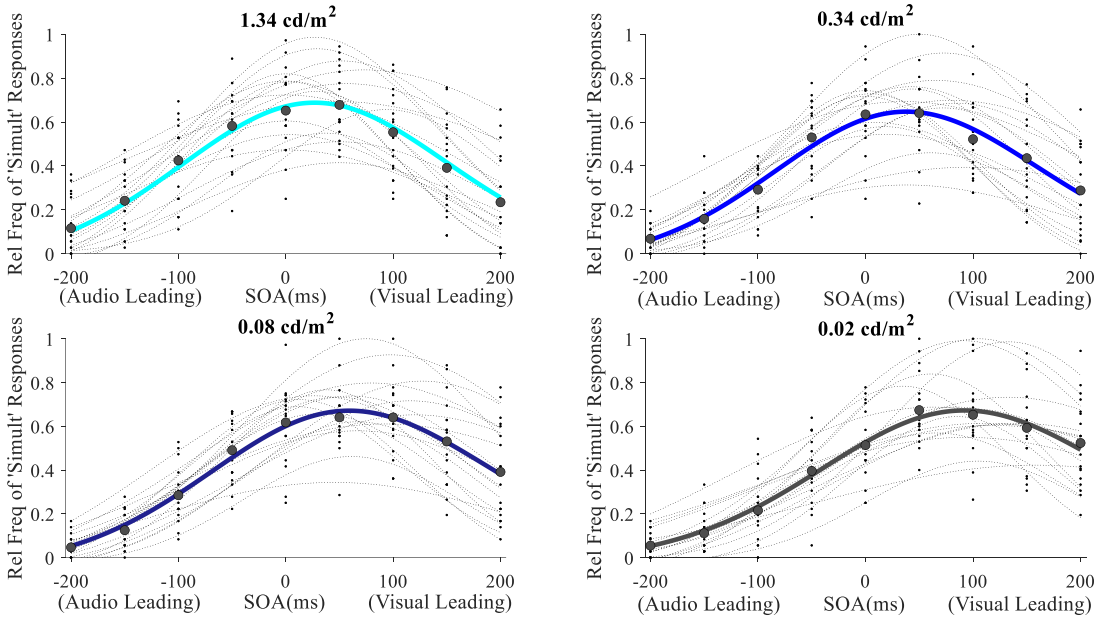


*Figure 5.4a. Simultaneity judgement data. A gaussian function of the form f(x) = a + b . exp*(-0.5 * ((*x* - *c*) / *d*)$^2$) was fitted (dotted lines) to each observers relative frequency (small dots). *Thick coloured lines represent the simulated curves using the mean parameters, and the larger circles 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).*

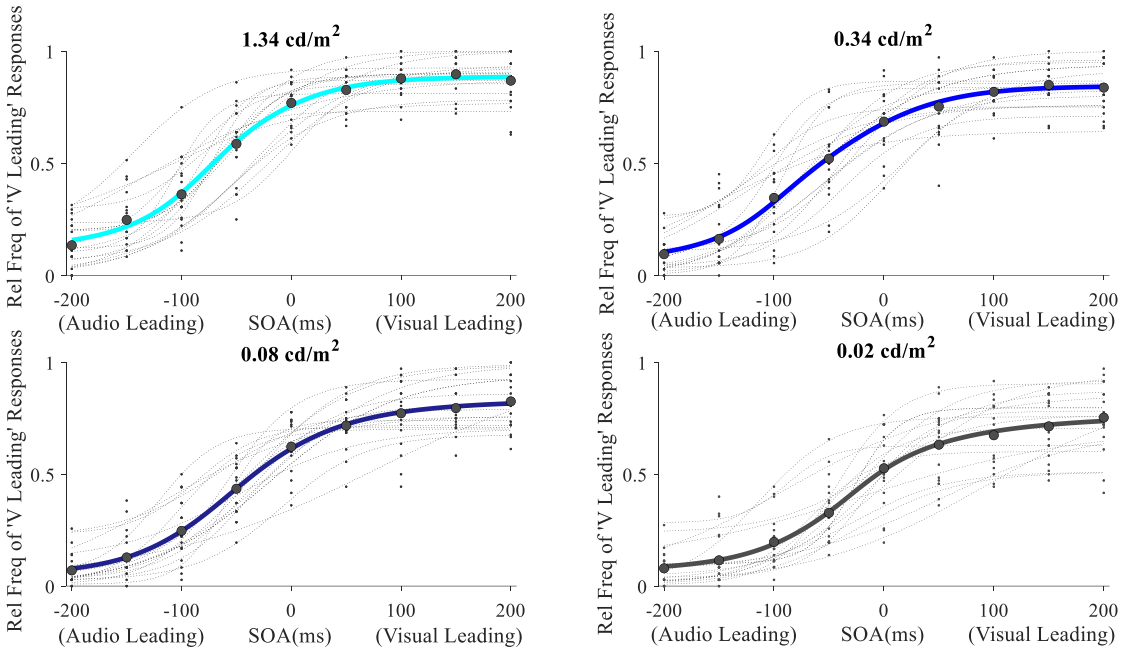


*Figure 5.4b. Temporal order judgement data. A sigmoidal Boltzman function of the form f(x) = (a - b) / (1 + exp((x - c) / d)) + b* was fitted (dotted lines) to each observers relative frequency (small dots). *Thick coloured lines represent the simulated curves using the mean parameters, and the larger 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 5.5 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.

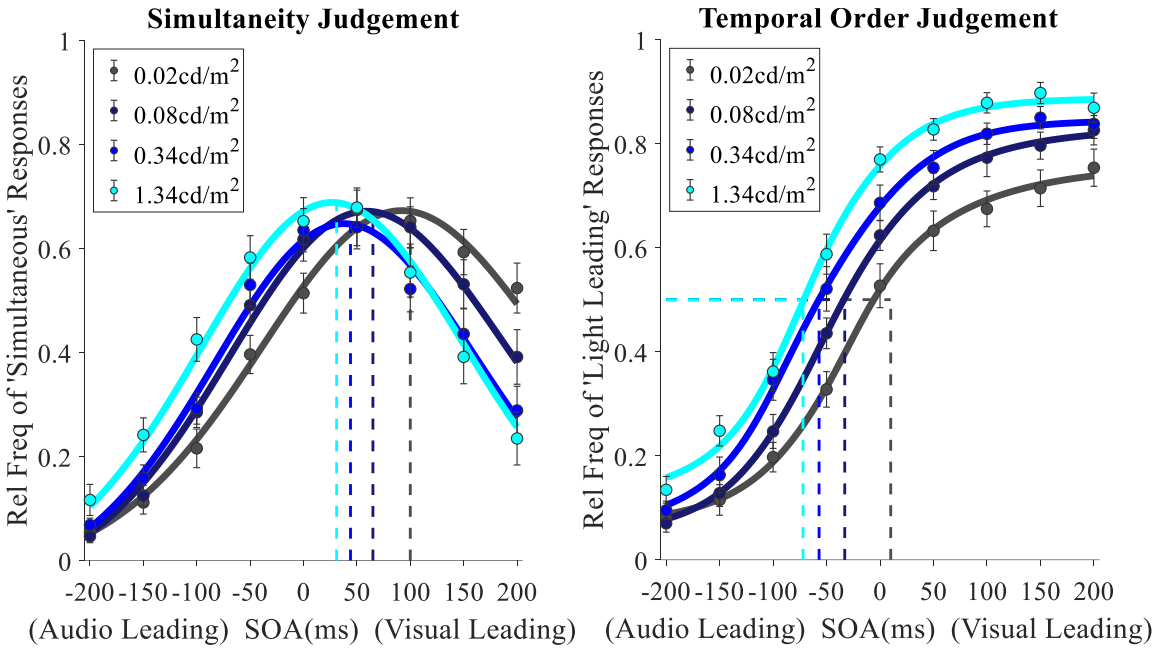


*Figure 5.5. Mean fits (solid lines) for the SJ (left) and TOJ (right) task. Filled in circles represent mean relative frequencies and error bars represent the standard error of the mean. Dashed lines represent the mean PSS across all participants at each of the 4 intensities. The mean PSS in the SJ 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$^2$).*

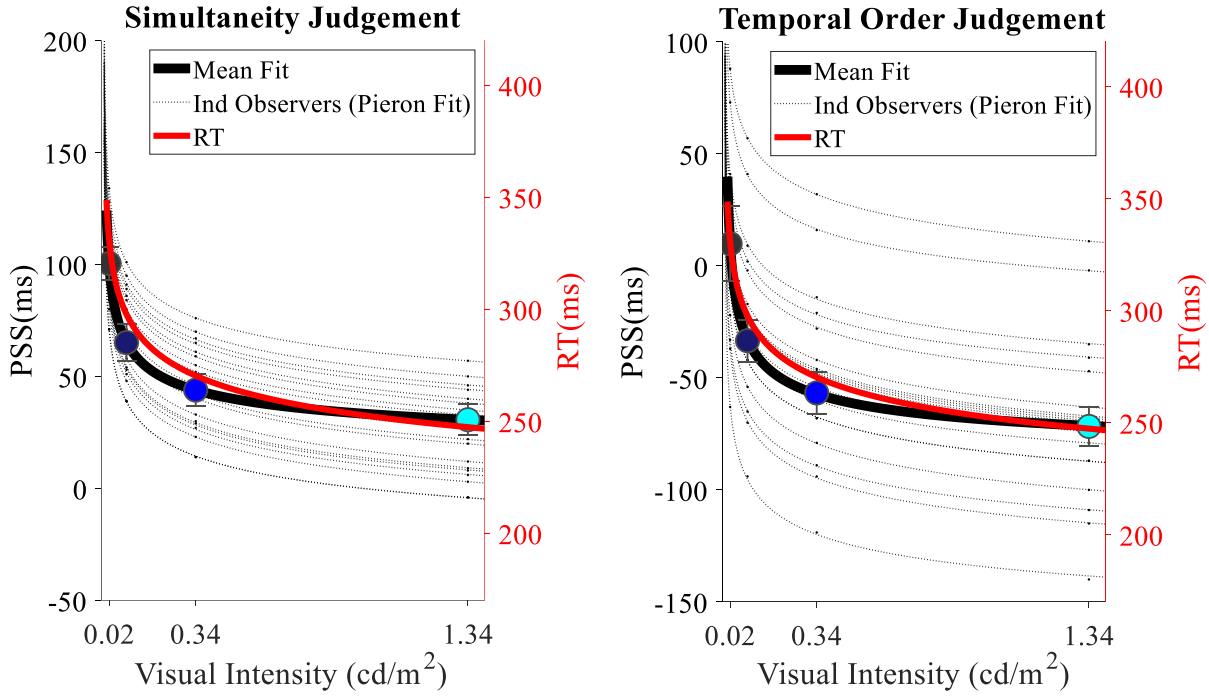


*Figure 5.6. A Piéron function* ( RT (I) = a.(I$^{-t}$) + $t_0$ ) *fitted to the individual PSS estimates (left y-axis) for the SJ (left) and TOJ (right) task as a function of intensity. The Piéron fit for the simple RTs is overlaid (right y-axis). The mean best fitting parameters (with SE) are as follows 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 5.6) 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).

### 5.6.2. Model evaluations

To compare the goodness of fits of the four models (Table 5.1, Section 5.4.2) 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):

$$AIC_c = n[\ln(RSS/n)] + 2k + [2k(k + 1)/(n - k - 1)] \quad (4)$$

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 5.C.1 (Appendix C) and Figure 5.8. 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 5.7 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.

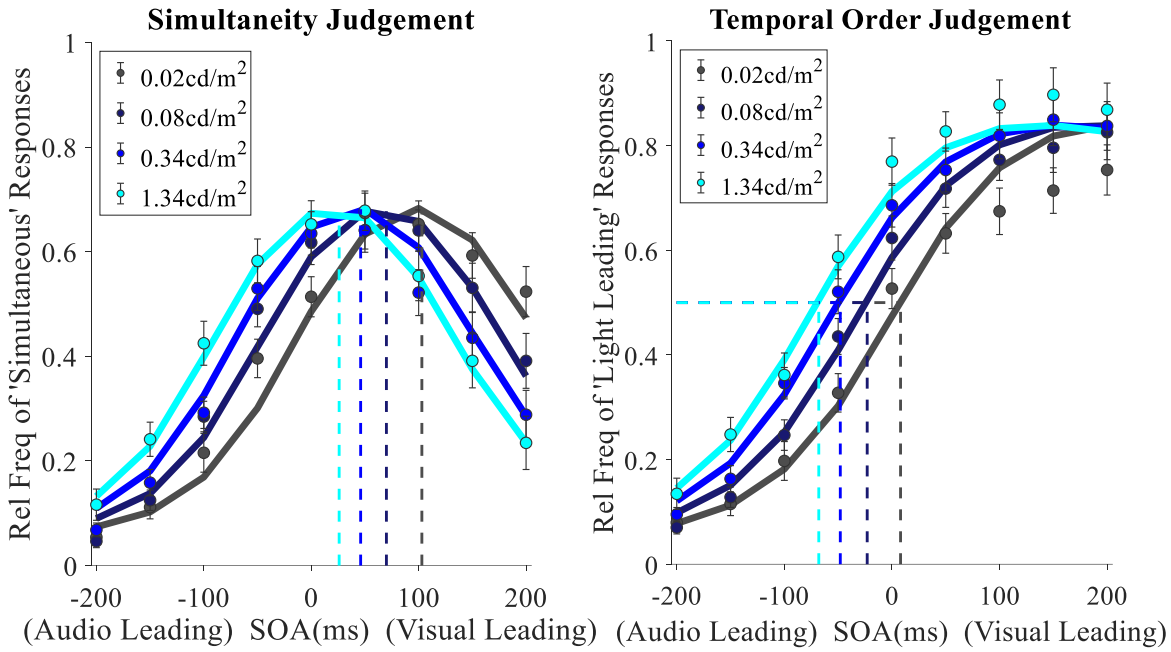


*Figure 5.7: Behavioural data fitted using model 3. The solid lines represent the data fitted using the mean parameters, and the dashed lines represents the mean PSS when model variant 3 was fitted individually to each participant. The mean PSSs are 37ms (1.34cd/m²), 45ms (0.34cd/m²), 61ms (0.08cd/m²) and 101ms (0.02cd/m²) in the SJ task and -55ms (1.34cd/m²), -47ms (0.34cd/m²), -31ms (0.08cd/m²) and 8ms (0.02cd/m²) in the TOJ task.*

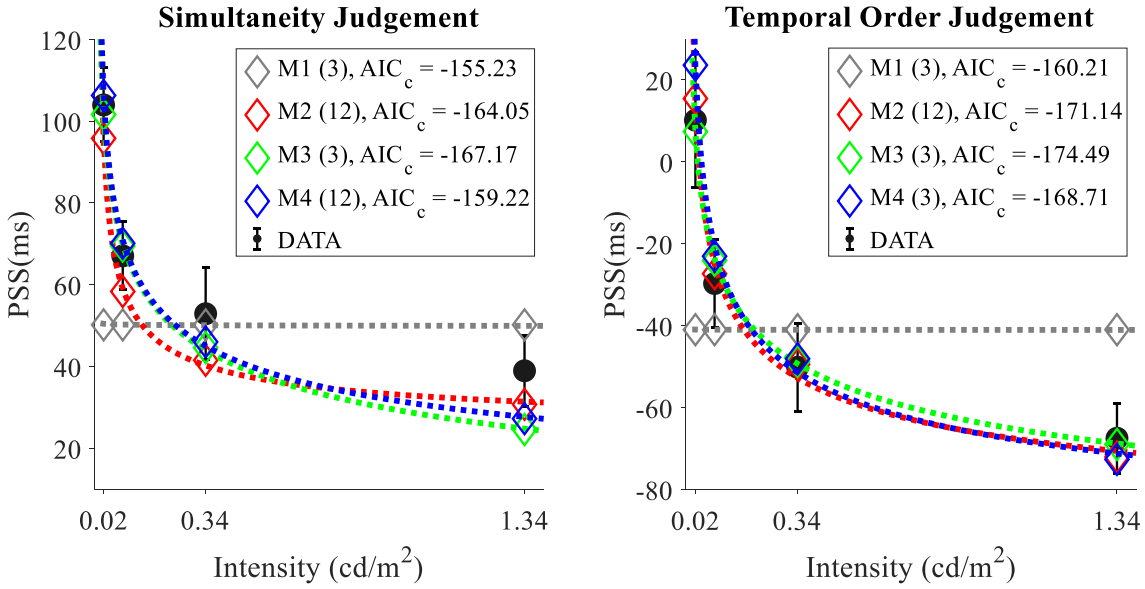


*Figure 5.8. Comparison of the individual PSS estimates (+-SE for the SJ (left) and TOJ (right) tasks. Each colour refers to a different model prediction. Figure legend values in parentheses represent the number of fitted parameters for each model.*

The predicted PSS values for all models (diamonds) are shown in Figure 5.8 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).

### 5.6.3. Correlation between the PSS obtained in the SJ and TOJ tasks

Finally, the PSS values obtained in the SJ and TOJ task (Figure 5.9 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 5.9 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 5.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.

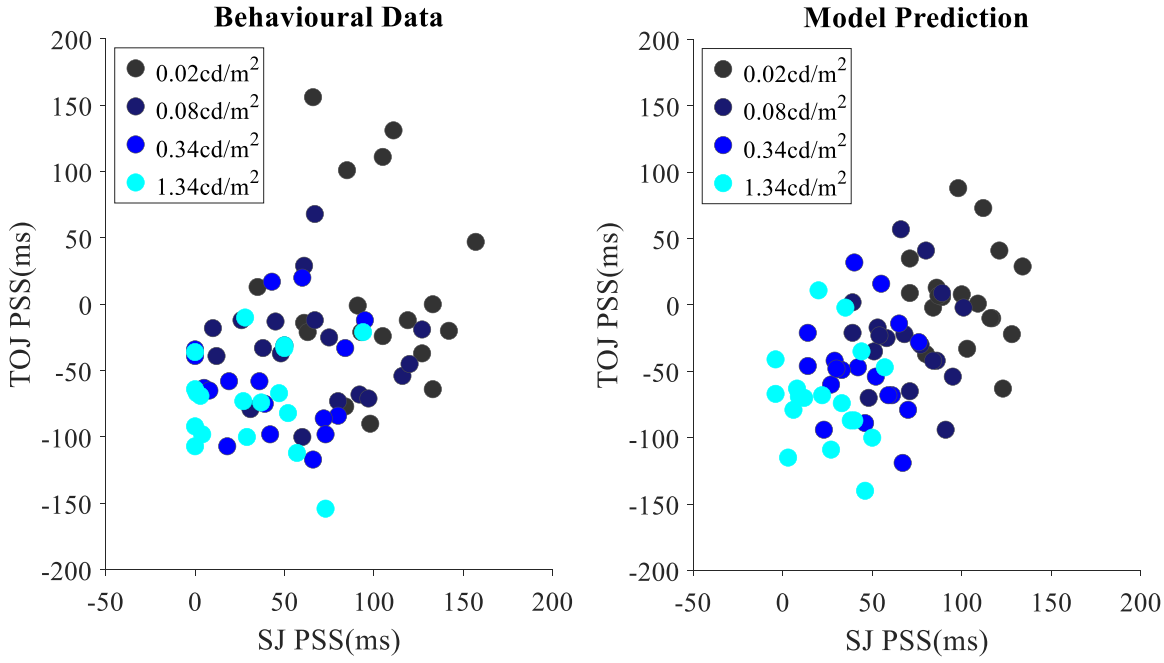


*Figure 5.9. Scatter plot showing the lack of correlation between the SJ and TOJ PSS values.*

## 5.7. 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.

### 5.7.1. Perceived audio-visual simultaneity is affected by intensity

The PSS values for the SJ task are generally shifted towards visual-leading (Figure 5.5) 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 et al., 2009; Stevenson, Zemtsov, & Wallace, 2012b; 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 5.6). This is consistent with a common mechanism underlying the perceived simultaneity in both tasks and will be explored in the next section.

### 5.7.2. An early intensity-dependent processing delay predicts perceived simultaneity

We tested four different variants of the MMCD model (Figure 5.1), either with or without an early nonlinearity which delays the signal (see Table 5.1), and the weights of the Lag and

Correlation detector (Figure 5.1) 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 5.8 and Table 5.C.1) but Model 3 is more the parsimonious model since only 3 parameters are estimated instead of 12. Figure 5.7 shows the predicted psychometric functions for model 3. Comparison with Figure 5.6 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 5.C.1; Figure 5.8). 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 5.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 5.D.1, left panel). We observe a similar dependence on intensity in the SJ and TOJ tasks, as shown in Figure 5.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 et al., 2007).

### 5.7.3. 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 5.8). 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.

### 5.7.4. 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.

## 5.8. Chapter summary

The results of the current chapter were consistent with the idea that the previously reported effect of visual intensity on the PSS (e.g., Leone & McCourt, 2015) is driven by intensity-dependent processing latencies that are common across both the SJ and TOJ tasks. Importantly, the data reported in both the current and previous chapter show the PSS of the two tasks to be uncorrelated, which the MCD model predicts by assuming two internal processing stages, which are differentially weighted at the decisional stage. The following chapter builds upon this using EEG to assess the timing of any differences in the neurocognitive processes underlying the two tasks. If the ERP data is consistent with the MCD model, then it would be expected that the early processing of the SJ and TOJ tasks would be identical, and that differences would only be observed later in the ERP, indicative

of differences at the decisional stage (e.g., Ahmadi, McDevitt, Silver, & Mednick, 2018; Gohil, Hahne, & Beste, 2016). Furthermore, it investigates whether the perception of simultaneity shifts to compensate for the intensity-dependent processing latencies of a preceding trial, akin to the previously observed rapid recalibration to the SOA of a preceding trial (Van der Burg, Alais & Cass, 2013; Simon, Noel & Wallace, 2017; Noel, Niear, Van der Burg & Wallace, 2016; Van der Burg, Orchard-Mills & Alais, 2015), which could perhaps explain the previously reported discrepancy in the effect of visual intensity on the PSS and RTs, in experiments where visual intensity was only manipulated across blocks (Harris et al., 2010; Jaśkowski, 1992; Menendez & Lit, 1983).

### 5.9. Appendix 5.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 $\tau$ 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 $MCD_{LAG}$ .

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

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

The $MCD_{CORR}$ and $MCD_{LAG}$ outputs are then averaged over a 5 second window, and the resulting mean is then normalised:

$$\overline{MCD_{CORR}} = \frac{mean[MCD_{CORR}(t)]}{\sum[S_A(t) * f_A(t)] + \sum[S_V(t) * f_V(t)]} \quad (A.6)$$

$$\overline{\text{MCD}_{\text{LAG}}} = \frac{mean[\text{MCD}_{\text{LAG}}(t)]}{\sum[S_A(t) * f_A(t)] + \sum[S_V(t) * f_V(t)]} \quad \text{(A.7)}$$

The $\overline{\text{MCD}_{\text{CORR}}}$ and $\overline{\text{MCD}_{\text{LAG}}}$ outputs are then mapped into behaviourally obtained psychometric data via a probit link function:

$$F(Y) = Y' = \beta_0 + \beta_{\text{CORR}} * \overline{\text{MCD}_{\text{CORR}}} + \beta_{\text{LAG}} * \overline{\text{MCD}_{\text{LAG}}} \quad (A.8)$$

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

### 5.9.1. Appendix 5.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
```

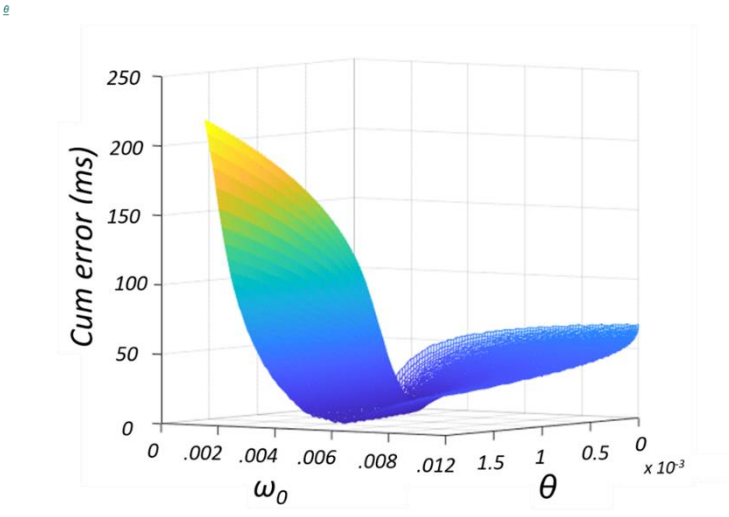


*Figure 5.B.1 – The effect of varying the cut-off and threshold on the cumulative error.*

### 5.9.2. Appendix 5.C.

| Model no. | Pre-Processing? | Task | No. of Free Parameters | Visual Intensity (cd/m²) | $\beta_0$ | $\beta_{CORR}$ | $\beta_{LAG}$ | PSS | RSS | $AIC_c$ | $\Delta AIC_c$ | 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 5.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 ($\overline{MCD_{CORR}}$) in the SJ task, and the lag detector ($\overline{MCD_{LAG}}$) in the TOJ task are evident.*

### 1.1.1. Appendix 5.D.

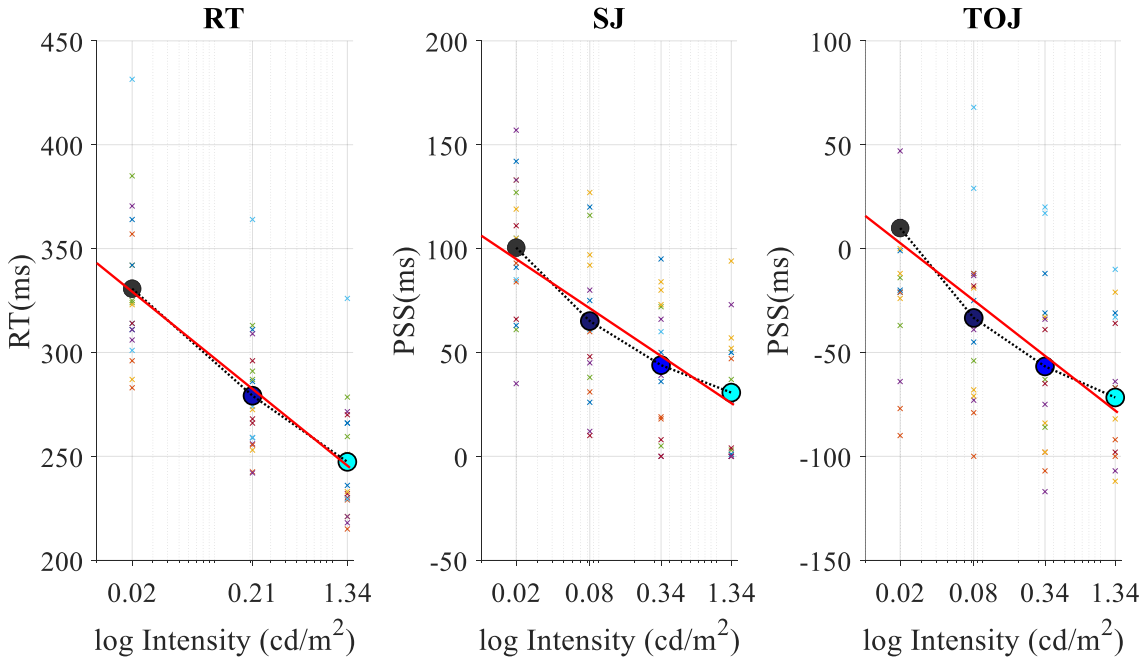


*Figure* 5.*D.1- Reaction times and PSS values (for SJ and TOJ) are plotted as a function of log intensity. Individual observer data (x) and the mean data (o) are shown, together with the best linear-log fit (red line). The slope for the reaction times (-19.9) is steeper than for SJ (-14.9) and similar to TOJ (-18.1).*

# 6. The rapid recalibration of perceived simultaneity, based on the visual intensity of a preceding trial

## 6.1. Chapter Introduction

Previous research has highlighted a process, whereby the perception of simultaneity is rapidly recalibrated to account for differences in the physical temporal offsets of bimodal stimuli (Van der Burg et al., 2013, 2018). Of course, it is rare that the relative onsets of audio-visual events rapidly vary in the real-world; since chapter 5 highlighted the role of intensity-dependent processing latencies in judgements of perceived timing, the current chapter aimed to investigate whether the PSS rapidly recalibrates to intensity-dependent variations in *central arrival times*, as opposed to solely in response to physical temporal offsets.

The current chapter therefore reports the results of a behavioural experiment (experiment 1) and an ERP experiment (experiment 2), which aimed to investigate whether the PSS rapidly recalibrates based on the visual intensity of a preceding audio-visual stimulus. Since the experiments in chapter 4 and chapter 5 replicated past findings which showed the PSSs estimated using the SJ and TOJ tasks to be uncorrelated, experiment 2 also aimed to investigate the spatiotemporal differences in the neural processes underlying the SJ and the TOJ tasks.

## 6.2. Abstract

A vital heuristic used when making judgements on whether audio-visual signals arise from the same event, is the temporal coincidence of the respective signals (Burr et al., 2009; Diederich & Colonius, 2015). Previous research has highlighted a process, whereby the perception of simultaneity rapidly recalibrates to account for differences in the physical offsets of stimuli (Van der Burg et al., 2013, 2018). The current paper investigated whether rapid recalibration also occurs in response to differences in central arrival latencies, driven by visual-intensity-dependent processing times.

In separate behavioural and ERP experiments, observers completed a TOJ and SJ task, and responded to audio-visual trials that were preceded by other audio-visual trials with either a bright or dim visual stimulus. It was found that the PSS shifted, due to the visual intensity of the preceding stimulus, in the TOJ, but not SJ task. The EEG data revealed a significant effect of preceding visual intensity on ERP across both tasks, and that this was evident only late (>300ms) in the ERP. Importantly, these differences highly overlap the time points and electrodes previously implicated in the effect of preceding *leading modality* on the ERP (Simon, Noel, Wallace, et al., 2017). It was therefore concluded that rapid recalibration to preceding intensity is driven by a decisional process that is consistent with that which underlies rapid recalibration to preceding leading modality.

## 6.3. Introduction

Neural processing latencies of sensory signals depend on several factors, including stimulus modality and intensity (Jaśkowski et al., 1995; Pöppel, et al., 1990; Ulrich et al., 1998). Auditory signals typically reach the respective sensory cortex at shorter latencies than visual stimuli (Raij et al., 2010), however an intensity dependency is evident for both modalities, with shorter processing latencies (Carrillo-de-la-Pena et al., 1999; Jaskowski et al., 1994; Plainis et al., 2013) and detection times (Pöppel et al., 1990; Ulrich et al., 1998; Vaughan et al., 1966) for higher intensity stimuli. One of the consequences of this intensity dependence, is that the perceived simultaneity of an audio-visual event depends on the relative intensities of the unimodal signals (Horsfall, Wuerger, & Meyer, 2021; Leone & McCourt, 2015; Roufs, 1963).

### 6.3.1. Recalibration of perceived simultaneity

It has been argued that humans recalibrate their audio-visual integration mechanisms, to deal with the different arrival times of auditory and visual signals (Fujisaki et al., 2004), thereby achieving a perception of simultaneity which is, to a large extent, independent of stimulus-related processing properties. This recalibration has been shown to occur after the prolonged adaptation to asynchronous audio-visual stimuli (Di Luca et al., 2009; Fujisaki et al., 2004; Navarra et al., 2009; Vroomen et al., 2004), though it is uncertain as to whether this occurs due to changes in early processing latencies (Di Luca et al., 2009; Navarra et al., 2009) or shifts in decisional criteria (Yarrow et al., 2011). More recently, it has been shown that recalibration also occurs on a dynamic, trial-to-trial basis (Noel et al., 2016; Simon, Noel, & Wallace, 2017; Van der Burg et al., 2013; Van der Burg, Orchard-Mills, & Alais, 2014). This putative rapid recalibration mechanism has been argued to have an important role in keeping our multisensorial perception stable in dynamically changing environments (Roseboom, 2019). The primary aim of the current paper is to investigate

whether this mechanism also compensates for differences in processing latencies that are driven by stimulus intensity.

### 6.3.2. Proposed intensity-dependent compensatory mechanism

An ideal audio-visual integrator should be independent of intensity, as intensity variations in processing latencies do not reflect the physical timing properties of stimuli in real-world scenarios. Since intensity differences are mapped into neural time delays early in the processing pathway, we hypothesise that rapid recalibration (based on the intensity information in the preceding trial) could play a role in minimising the intensity dependence of perceived audio-visual simultaneity. In other words, we hypothesise that recalibration is modulated by the relative, intensity-dependent central arrival times of audio-visual signals, rather than simply the physical offset of these signals. Since low intensity visual signals reach cortical processing areas at longer latencies, we predict that this processing delay will be compensated for as if it were a physical stimulus offset (e.g., a dim visual flash presented simultaneously with an auditory 'beep' could be treated as an auditory-leading stimulus, whereas a bright visual stimulus could be treated as visual-leading), resulting in a PSS shift.

Figure 6.1D displays the quantitative prediction. The blue dashed line represents a regression line fitted to the predicted PSS, as a function of preceding SOA; the magnitude of this predicted shift was based on the data from Simon, Noel and Wallace (2017). The predicted effect of preceding intensity (cyan and grey dashed lines), as a function of preceding SOA, is based upon previously reported PSSs across varying visual intensities (Horsfall et al., 2021). Since visual processing latencies are intensity dependent (Lines et al., 1984), visual signals in bright trials (figure 6.1B) will reach multisensory areas of the brain at shorter latencies than dim signals (figure 6.1C). The offset in these intensity-dependent central arrival times will then be recalibrated to in the subsequent trial (figure 6.1D), as if

they were physical temporal offsets. Again, the magnitude of this shift was predicted, based upon the effect of preceding SOA on PSS (Simon et al., 2017). It was therefore predicted that the PSS would shift towards the visual-leading SOAs following bright stimuli, and towards the auditory-leading SOAs following dim stimuli.

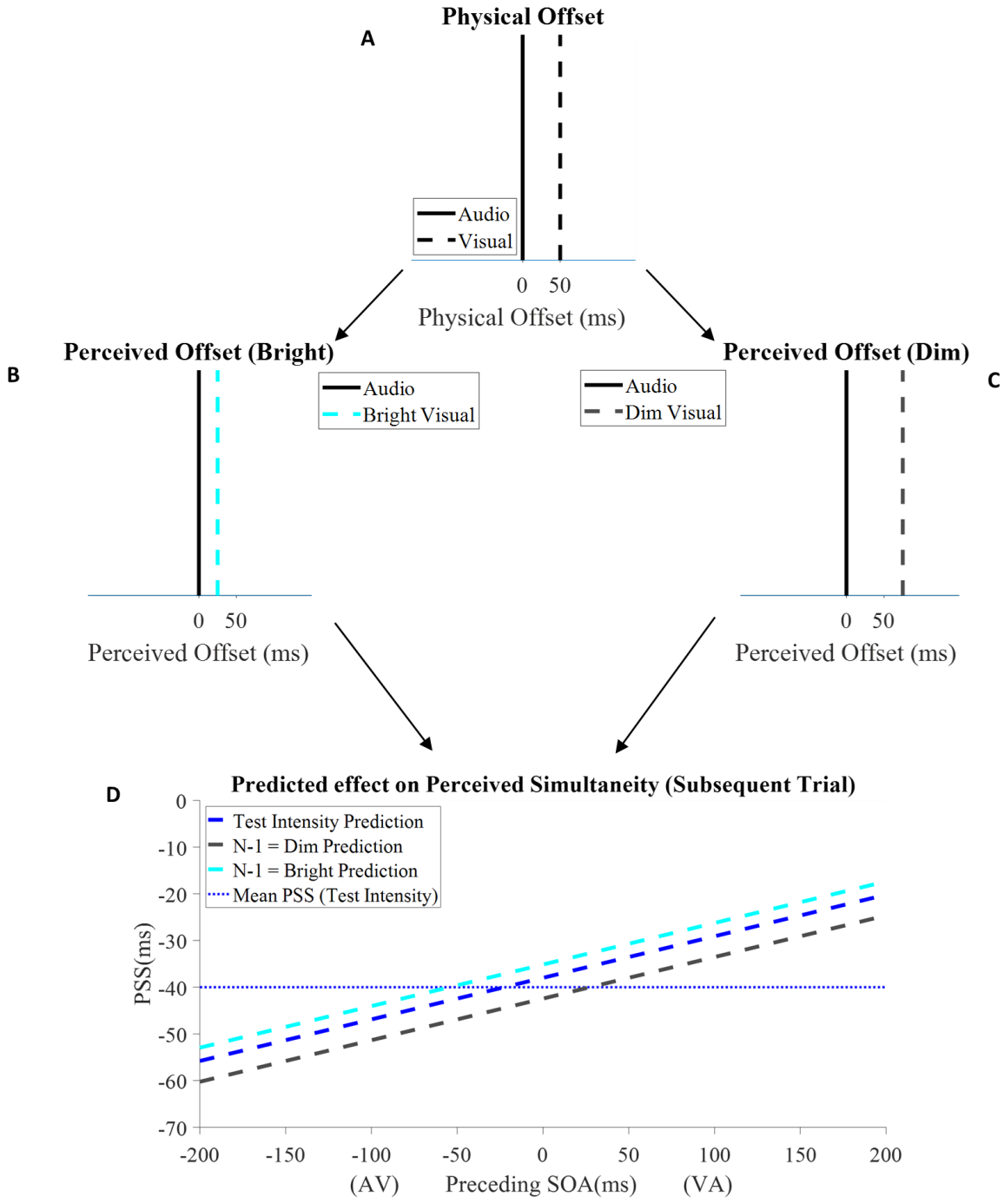


*Figure 6.1 - Prediction of how the PSS would shift, to compensate for audio-visual stimuli with different visual intensities. 1A, 1B and 1C reflect schematic diagrams of how the perception of a single physical stimulus offset (1A) can be affected by visual intensity-dependent processing latencies (1B and 1C). In this example, an auditory beep is presented*

*50ms before a visual flash (1A). The bright visual stimulus (1B) is processed at a shorter latency than the auditory stimulus, and therefore the perceived (audio-leading) offset between the audio and visual signals is reduced in comparison to the physical offset. The dim visual stimulus (1C) has a longer processing latency than the auditory stimulus, and therefore the audio-leading offset is increased, relative to the physical offset. 1D displays a quantitative prediction of a PSS shift, based on the preceding trials visual intensity and physical offset. The horizontal dotted blue line represents a predicted PSS for the test intensity stimulus, which was predicted by taking the average of the PSS at the 0.08 and 0.34 cd/m2 intensities of the TOJ data from Horsfall et al., (2021). The blue dashed line represents a regression line fitted to the predicted PSS, as a function of preceding SOA. The cyan (N-1 = bright) and grey (N-1 = dim) dashed lines represent regression lines fitted to the predicted PSS shift, based on the intensity of the preceding trial. The magnitude of the preceding-intensity-dependent shift, was predicted using the PSSs reported in Horsfall et al., (2021): here, the difference between the highest intensity PSS and the predicted PSS for the test intensity was taken, as well as the difference between the lowest intensity PSS and the predicted PSS for the test intensity. These were used to infer the difference in visual processing latencies between the stimuli, and to predict the perceived offset at a given physical SOA and visual intensity. The point on the test intensity prediction (blue dashed line), corresponding to the predicted perceived offset, was then taken to predict the effect of preceding intensity and SOA on PSS.*

### 6.3.3. Time course of a proposed intensity dependent compensatory mechanism

An additional aim of the current study was to understand the time course of the neurocognitive processes related to the putative compensatory mechanism. Event-related potentials (ERPs) are ideally suited to investigating the temporal dynamics of neural processes, due to their high temporal resolution in the order of milliseconds. ERPs have been used extensively to investigate the time course of processing related to auditory-visual integration mechanisms (e.g., Harrison et al., 2015; Meyer, Harrison, & Wuerger, 2013; Molholm, Ritter, Javitt, & Foxe, 2004), and, importantly, in relation to multisensory temporal simultaneity (e.g., Simon et al., 2017). Rapid recalibration to the leading modality of the preceding stimulus has been shown to modulate only later (i.e. > 300ms) portions of the ERP waveform (Simon et al., 2017), indicative of decisional processes (O'Connell, Dockree, & Kelly, 2012; Twomey, Murphy, Kelly, & O'Connell, 2015). Hence, we predicted that rapid recalibration to preceding intensity will likewise modulate only the later portion of the ERP waveform.

### 6.3.4. Task specific processes

Traditionally, simultaneity judgment (SJ) and temporal order judgement (TOJ) tasks are used, either separately or in combination, to study perceived simultaneity. However, the tasks have been shown to estimate differential effects of rapid recalibration (Roseboom, 2019), and so we wished to investigate whether the hypothesised effect of preceding intensity is consistent across the tasks. Furthermore, the tasks often produce uncorrelated PSS estimates (e.g., Love et al., 2013; Van Eijk et al., 2008), leading to the argument that they draw upon separate neural mechanisms (Jaśkowski, 1991; Stelmach & Herdman, 1991; Love et al., 2013, Zampini, 2003). A secondary aim of the experiment was to investigate when in the processing stream these tasks differ. A recent processing model showed that these differences can be explained by divergent decisional processes (Parise & Ernst, 2016), and so it was predicted that this will be reflected by late (> 300ms) differences in the ERP.

### 6.3.5. The current study (overview and purpose)

The main aims of the study were as follows: A) to investigate whether rapid recalibration is modulated by the relative, intensity-dependent central arrival times of audio-visual signals B) to elucidate the neural time course of this proposed compensatory mechanism, C) to investigate whether this proposed shift in perceived simultaneity is reflected by changes in early processing latencies, as evident in ERP latencies and simple reaction times (RTs). A secondary aim of this paper was to assess the difference in the neurocognitive mechanisms underlying the two tasks (SJ and TOJ) which are commonly used to assess the perception of simultaneity.

In a behavioural experiment (experiment 1), audio-visual stimuli of three visual intensities and one auditory intensity were presented across a range of stimulus onset asynchronies

(SOAs). 21 participants completed the audio-visual simultaneity judgement (SJ) and temporal order judgement (TOJ) tasks, which are commonly used to estimate an observer's point of subjective simultaneity (PSS) in milliseconds, alongside a simple reaction time (RT) task. A separate sample of 20 participants completed the EEG experiment (experiment 2), where they completed the SJ and TOJ tasks, with identical stimulus intensities to experiment 1.

## 6.4. Experiment 1: Behavioural study

### 6.4.1. Method

#### 6.4.1.1. Participants

The behavioural study involved 21 participants, aged 19-49 (mean = 27.10, SD = 7.80, median = 25), 9 of which were female and 13 of which were naïve to the SJ and TOJ tasks. All observers reported normal or corrected-to-normal vision and hearing, and no neurological disorders.

#### 6.4.1.2. Design

A repeated measures design was used. Participants completed the audio-visual simultaneity judgement task, temporal order judgement task and three variants of a simple reaction time task on three separate days. The order of the TOJ and SJ tasks were counterbalanced, and the RT task was always completed second.

#### 6.4.1.3. Apparatus

Participants were seated in a soundproof booth (IAC, Winchester, UK), 113cm from a LED and speaker that were held at roughly eye level in an adjustable clamp. A Tucker Davies RP2.1 real-time processor (TDT technologies, Alachua, FL) was used to generate the visual and auditory stimuli. A single 'Xenta M-219 Notebook speaker' produced the auditory stimuli. This was located 1.62° below a 5mm white LED. A custom-built button box,

consisting of two wide, shallow buttons was used to record the participant's responses. MATLAB was used on a PC located outside of the soundproof booth to signal the Tucker Davies system and record responses.

#### 6.4.1.4. Preliminary Threshold Estimation

The visual and auditory stimuli were chosen by estimating the visual and auditory thresholds of 8 participants (age range: 22-43, mean = 27.00, SD = 2.39, median = 24.50) in a pilot experiment. After 15 minutes of dark adaptation, participants completed two separate two-alternative forced choice (2AFC) tasks, consisting of a visual threshold estimation task, followed by an auditory threshold estimation task. The threshold estimation used a QUEST (Watson & Pelli, 1983) procedure as this is more efficient than the typical threshold estimation involving sequential testing (Watson & Pelli, 1983). We implemented two interleaved staircases consisting of 40 trials each. 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 occurred after one of two clearly detectable visual flashes. The final threshold was estimated to be the minimum intensity at which the participant could detect the stimulus with ~75% accuracy. The mean visual threshold was 0.005 cd/$m^2$ and the mean auditory threshold was calculated as 20.54db.

#### 6.4.1.5. Stimuli

The stimuli consisted of a 100ms visual flash and a 100ms auditory beep. The visual flash was presented at three intensities (0.02 cd/$m^2$, 0.16 cd/$m^2$ and 1.34cd/$m^2$) which were 6db,

15db and 24db above the previously estimated mean visual threshold. The auditory beep had a fixed frequency of 1000hz, and was presented with a flat amplitude envelope. The intensity of the auditory stimulus was 15dB above the previously estimated threshold (see section 6.4.1.4) in an attempt to make sure the intensity was comparable to that of the visual stimuli. The inter-stimulus interval was 1500ms, plus a random value between 0-2000ms.

#### 6.4.1.6. Behavioural trials

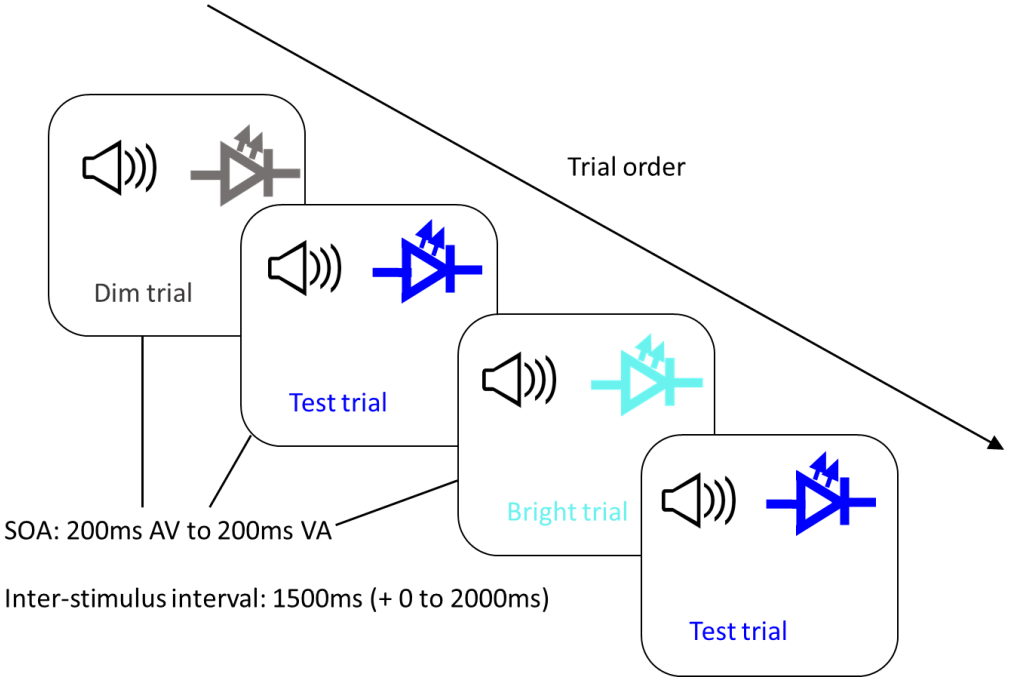


*Figure 6.2 - Depiction of an example trial order. Each trial consisted of an audio-visual (speaker and LED symbols) stimulus, that was either audio-leading, visual leading or simultaneous. Trials were presented so that every other trial was at the 'test' intensity, and that these test intensity trials were always preceded by either a dim (grey), bright (cyan) or filler (see below) trial. Observers were asked to respond to every trial.*

Stimuli were identical across the SJ, TOJ and RT tasks. The audio-visual stimuli were presented at 7 SOAs (-200ms, -100ms, -50ms, 0ms, 50ms, 100ms, 200ms). The 3 intensities were presented so that every second trial (aside from the filler trials) was of the middle (0.16 cd/m$^2$) intensity, which we define as the 'test' intensity. Within each block, a trial at each of the 7 SOAs of the test intensity was preceded by a trial at each of the 7 SOAs of the

dim (0.02 cd/$m^2$) and bright (1.34cd/$m^2$) intensities, such that each of these SOAs were presented in random order. Therefore, there were 98 test trials in each block, which were preceded by 49 dim trials and 49 bright trials, plus 32 'filler' trials. These consisted of two audio-visual trials with random flash intensities (either bright or dim), each presented with a random offset between (-200 and 200ms), followed by a test intensity trial. These were included to ensure participants did not become aware of the pattern of stimulus presentations. All filler trials, alongside the following test intensity trials were not included in any analyses. Both the SJ and TOJ tasks consisted of 7 blocks each. Over the course of these 7 blocks were a total of 98 trials at each SOA of the test intensity and 49 trials at each SOA in the dim and bright conditions, totalling 1596 trials in each task (including pairs of filler trials). In the RT task, participants completed 4 blocks of trials which are identical to those presented in the SJ task, summing to 488 trials in total.

#### 6.4.1.7. Data fitting

To estimate the PSS in the TOJ task, a sigmoidal curve (Boltzmann) was fitted to the relative frequency of 'light first' responses:

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

where *a* reflects the upper asymptote, and *b* reflects the lower asymptote. The parameter *d* represents the width of the curve, whereas *c* was the location of the curve. The SOA (in ms) corresponding to a 50% response rate of 'light first button presses was used as an estimate of the PSS.

For the SJ data, a Gaussian was fitted to the relative frequency of 'simultaneous' responses:

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

where *a* reflects the offset along the y-axis, *b,* the amplitude of the curve and *c* reflects the midpoint which was used as an estimate of the PSS. The parameter *d* reflects the width of

the curve. A nonlinear least-squares method (Function 'fit'; MATLAB2018) was used to find the best-fitting parameters for both the TOJ and SJ data.

#### 6.4.1.8. Data Analysis

A one-way repeated measures ANOVA was used to assess the effect of visual intensity on the PSS in the SJ and TOJ tasks. Here, a Greenhouse-Geisser correction was applied where necessary. The Pearson correlation coefficient was used to assess the correlation between PSSs recorded in the SJ and TOJ tasks. This was done separately at each of the three intensities.

The test intensity data of the SJ and TOJ task were split based on the SOA and visual intensity of the preceding trial, and the resultant relative frequencies (either of 'light first' or 'simultaneous' responses) were fitted using functions described in section 6.4.1.7. The PSSs of four participants were estimated to be outside of the testable range in at least one of the conditions of this analysis, and they were therefore removed from all analyses. Regression lines were fitted independently to each observer's PSSs, as a function of preceding SOA. This was done separately for trials preceded by dim or bright trials (i.e. split data), and for trials preceded by both bright and dim trials (i.e. all test intensity trials). Paired *t-tests* were used to compare the intercepts of the regression lines fitted to data preceded by dim or bright trials. One-tailed tests were used, as the prediction for the effect of preceding intensity was directional (see figure 6.1).

For the RT data, a 7 x 3 repeated-measures ANOVA was used to assess the effect of SOA (-200 – 200) and intensity (dim, test intensity, and bright) on reaction-times (Simon et al., 2018). A 7 x 2 x 2 repeated measures ANOVA was then used to assess the effect of SOA, preceding intensity, and preceding leading modality (audio-leading or visual-leading), on RTs at the test intensity. Greenhouse-Geisser corrections were applied where appropriate.

#### 6.4.1.9. Procedure

For the SJ task, participants were asked to hold the project box vertically, and to press the top button if they believed that the two stimuli occurred simultaneously and the bottom button if they occurred non-simultaneously. For the TOJ task participants were asked to hold the project box horizontally, and to press the left button if they perceived the flash to have occurred before the beep, and the right button if they perceived the beep first. Participants were requested to hold the button box with different orientations to create a clear distinction between the two tasks, in order to minimise the potential of the participant habitually performing one task, when they should have changed to the other. Participants were asked to respond to every trial and to use their best guess when unsure of the correct response. For the reaction-time task participants were instructed to respond as quickly as possible with a left button press whenever they perceived a light or sound (or both). Participants also completed a unimodal simple RT task, the data from which formed part of a separate experiment. Before both the initial SJ or TOJ block, participants completed two practice blocks of the upcoming task, consisting of 36 trials of the highest visual intensity, split across 3 SOAs (-200ms, 0ms and 200ms) in the SJ and RT task and 2 SOAs (-200ms and 200ms) in the TOJ task, in order to increase familiarity with the buttons. Participants would only progress to the main SJ and TOJ task if they achieved at least a 66.66% correct response rate in the respective practice session in one of their two attempts. An opportunity to ask questions was provided after both blocks. Participants were then given 15 minutes of dark adaptation before the experiment began, during which they were asked to relax. Short breaks were provided between each block. After the final (third) session was completed, the participants provided written answers to questions aiming to obtain basic participant information, along with a question on whether the participant had noticed any pattern in the trials/stimuli presented during the experiment. This final question was given with the intention of removing any participant who became

aware that the test intensity was always preceded by a dim or bright trial, though no participants reported this.

### 6.4.2. Results - Experiment 1

#### 6.4.2.1. Effect of intensity on perceived simultaneity

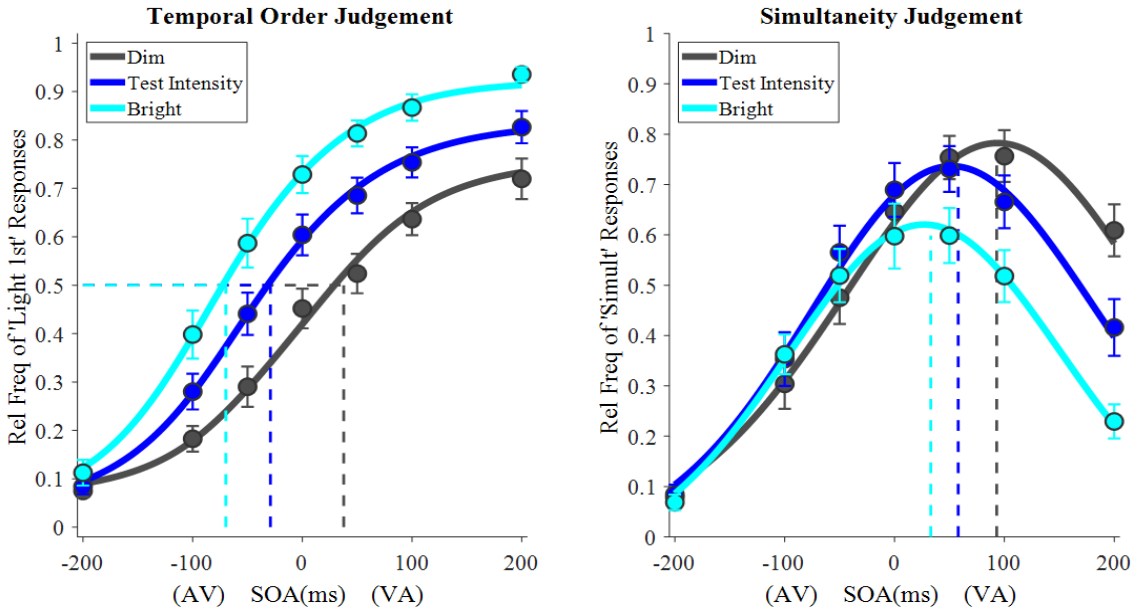


*Figure 6.3 – Mean fits across observers in the TOJ (left) and SJ (right) tasks at the three visual intensities. Data points represent mean relative frequencies of 'light first' (TOJ) or 'simultaneous' (SJ) responses (+/- SEM). Dashed lines represent the mean PSS at each intensity. The 'test' intensity data is later split based on the properties of the preceding stimulus. The mean estimated PSSs are as follows: TOJ dim: 38ms, TOJ test intensity: -29ms, TOJ bright: -70ms, SJ dim: 93ms, SJ test intensity: 58ms, SJ bright: 33ms.*

One of the main aims of our analysis was to investigate the effect of preceding intensity on PSS. However, we must first ascertain whether our intensity manipulation had a reliable effect on PSS across the two tasks. In the TOJ task, there was a significant main effect of intensity on PSS ($F(1.31, 20.99) = 34.51$, $p < .001$, $\eta p^2 = .68$). Bonferroni corrected post-hoc tests revealed significant differences between each of the intensities ($p < .01$), with higher PSSs (indicating the light had to come earlier, relative to the sound, in order for

simultaneity to be perceived) for dimmer stimuli. Likewise, in the SJ task, there was a significant main effect of intensity ($F(1.48, 23.71) = 49.00$, $p < .001$, $\eta p^2 = .75$), and post-hoc tests revealed significant differences between the PSSs at each intensity ($p < .01$), with higher PSSs for dimmer stimuli.

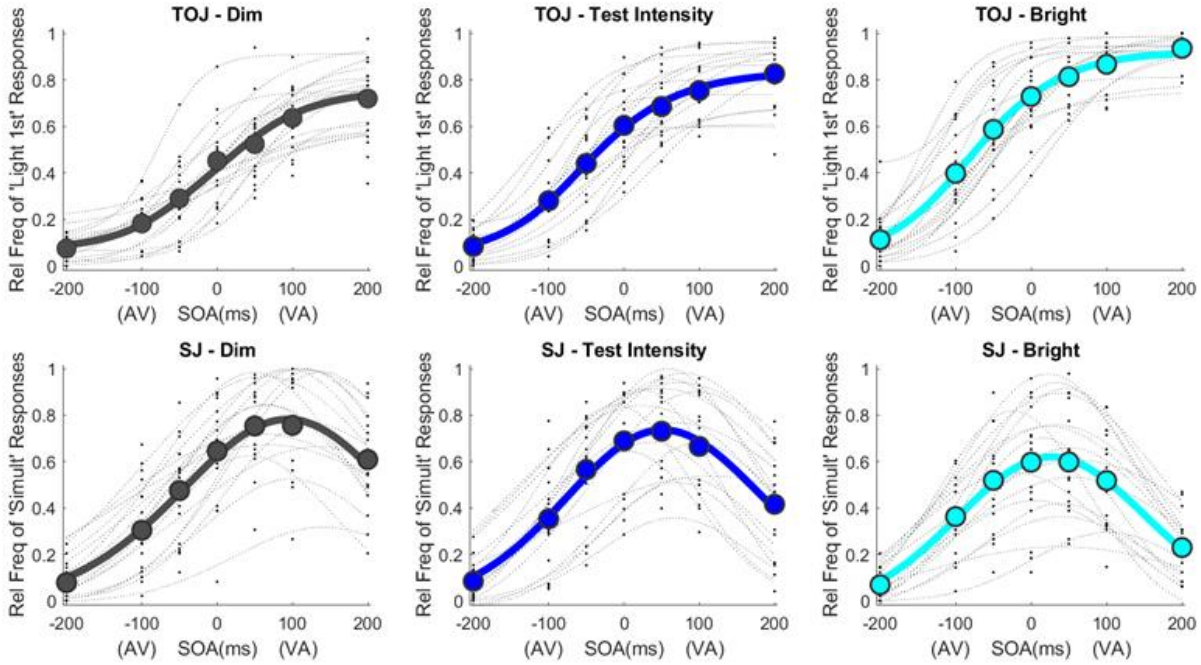


*Figure 6.4 - Individual fits for each observer (black dotted lines) in the TOJ (top row) and SJ (bottom row) tasks, at the dim (left column) test (middle column) and bright (right column) intensities. Each larger dot represents a relative frequency of 'light first' responses in the TOJ task, and 'simultaneous' responses in the SJ task. Thick lines represent mean fits.*

Previous research has shown the PSSs from the SJ and TOJ tasks to be uncorrelated (Basharat et al., 2018; Binder, 2015; Linares & Holcombe, 2014; Love et al., 2013; Van Eijk et al., 2008; Vatakis et al., 2008). The PSSs estimated from the current experiment are plotted in figure 6.5. While an effect of intensity is apparent across both tasks, the PSSs were uncorrelated at all three intensity levels (dim: $r(16) = .22$, $p = .397$; test: $r(16) = .44$, $p = .077$; bright: $r(16) = .28$, $p = .272$), confirming our data's consistency with past findings.

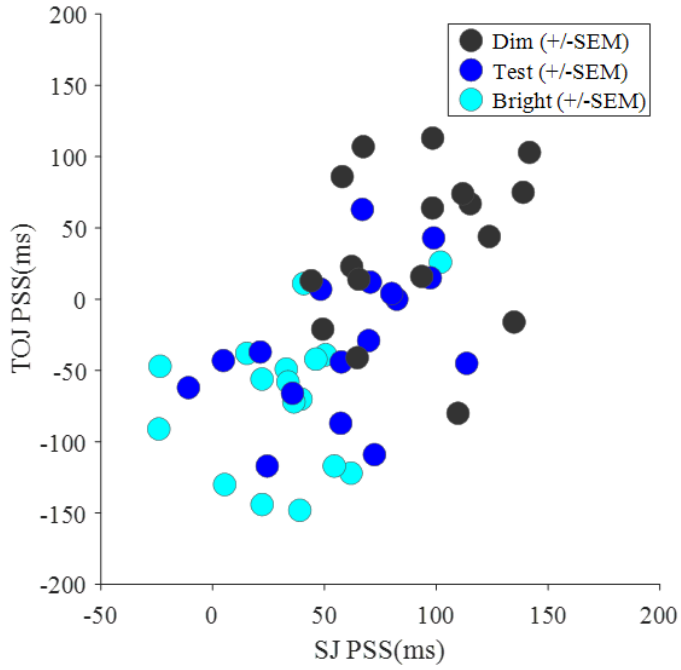


*Figure 6.5 – Scatterplot of the PSSs recorded from the SJ and TOJ task at the three intensities.*

#### 6.4.2.2. The effect of *preceding* intensity-driven perceived offsets on perceived simultaneity

The previous section shows a systematic effect of intensity on PSS and confirms that our behavioural data are consistent with past findings. The following section aims to assess the effect of the relative intensity and SOA of the preceding trial on the PSS. We therefore split our behavioural data, both by the visual intensity (n-1 = dim or n-1 = bright) and by the SOA (-200AV - 200VA) of the preceding trial. The data for each observer was individually plotted, based on the preceding intensity and SOA, to obtain the mean PSSs depicted in figure 6.6A and figure 6.7A.

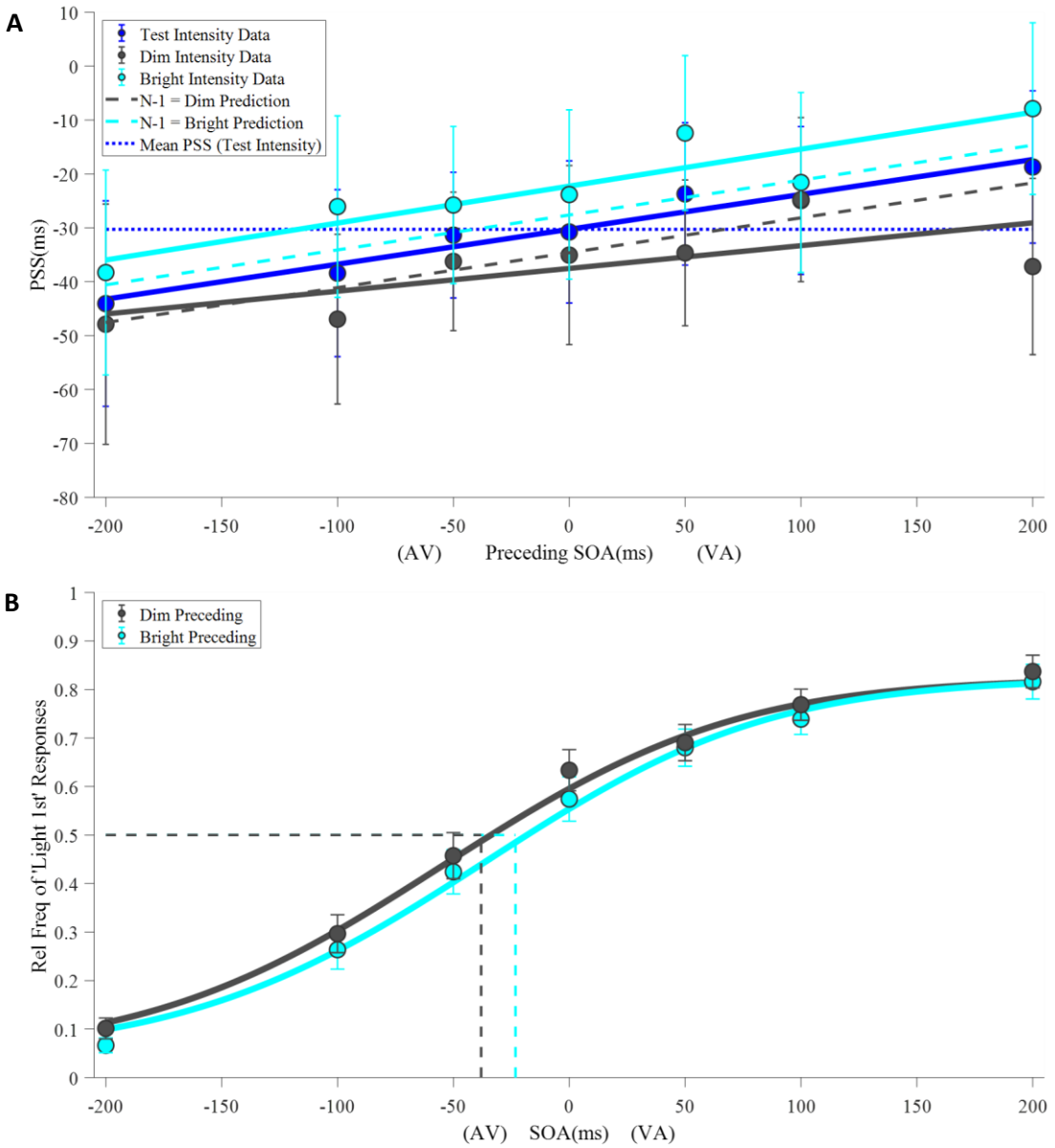


*Figure 6.6 – A) The effect of preceding intensity and SOA on the PSS in the TOJ task. Regression lines were fitted to the PSSs of trials preceded by dim (solid grey line) and bright (solid cyan line) intensity stimuli, as a function of preceding SOA across all observers. The mean difference between the intercepts was 15ms. The blue regression line reflects the medium intensity data that were not split based on preceding intensity, and the blue dotted line reflects the mean of the PSSs at the test intensity. Data points reflect mean PSSs (+/- SEM). The predicted shifts were based on the current TOJ data (see section 6.3.2 for how these predictions were made). B) The mean fits of the TOJ data, based only on whether the preceding trial was dim (grey line) or bright (cyan line) (i.e., combined over all preceding SOAs). Data points represent the mean relative frequency of pressing the 'light-first' button (+/- SEM). The dashed lines reflect the mean estimated PSSs, which are as follows: N-1 = dim: -38ms, N-1 = bright: -23.18ms.*

The TOJ data is plotted in figure 6.6; an effect of preceding SOA that is in line with the prediction (figure 6.1) was evident, with a larger magnitude shift for trials preceded by larger offsets. Here, a 25ms difference was observed between the mean PSS of trials preceded by a 200ms audio-leading stimulus, in comparison to a 200ms visual-leading stimulus. Importantly, an effect of preceding intensity was also evident, with a shift that is larger than predicted. A regression line was plotted to the PSSs of the trials, dependent on the intensity of the preceding trial (dim versus bright). The difference between the intercepts of the regression line fitted to the trials preceded by dim ($M$ = -37.53ms, $SD$ = 56.80) and bright ($M$ = -22.26ms, $SD$ = 55.29) trials was significant ($t$ (16) = 1.93, $p$ = .036, $d$ = 0.46) (one-tailed).

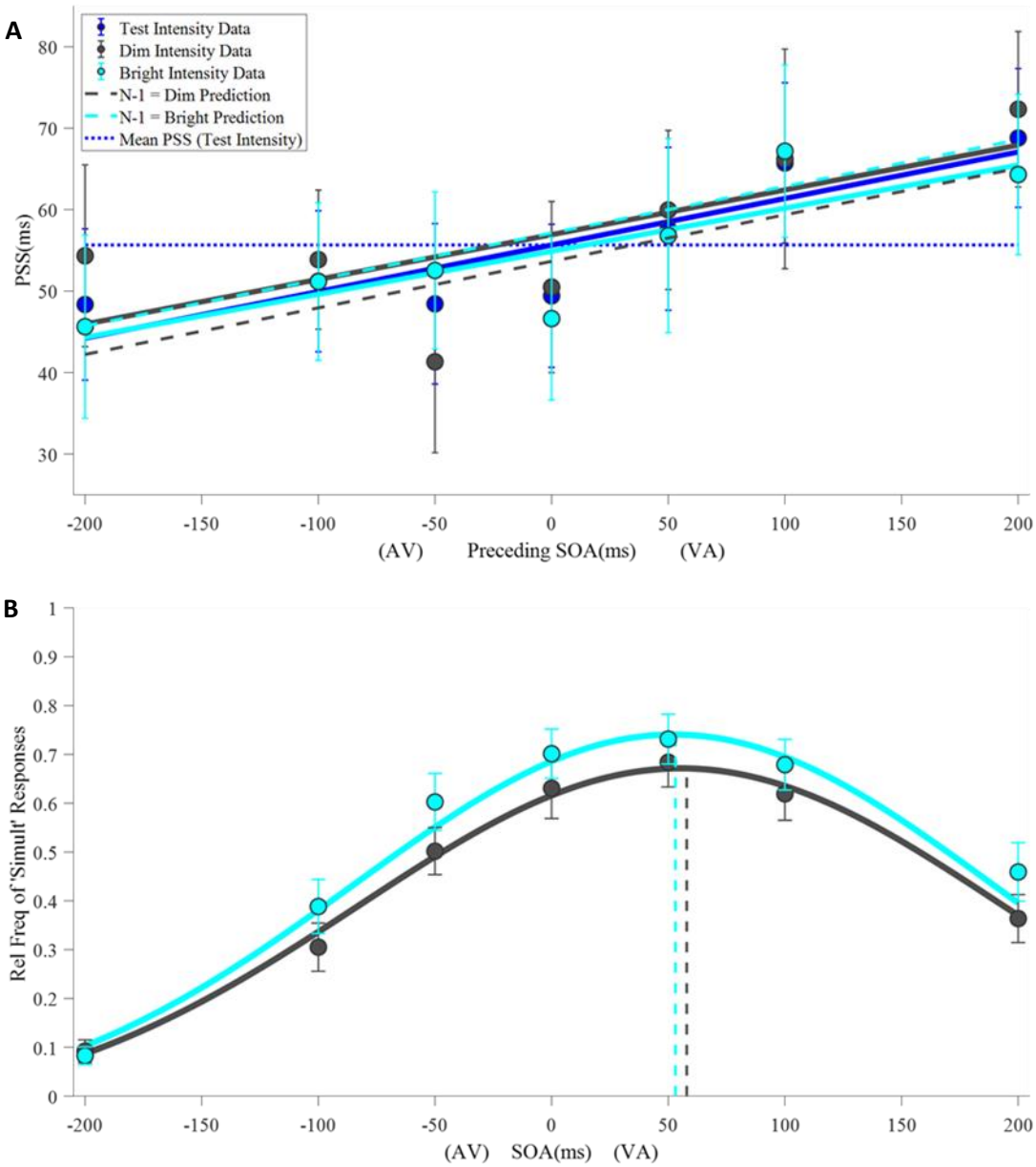


*Figure 6.7 – A) The effect of the preceding intensity and SOA on PSS in the SJ task (see legend of figure 6.6A for details). Note that the predicted shifts in the top subplot were based on the current SJ data. B) The mean SJ fits, based only on the intensity of the preceding trial (see legend of figure 6.6B for details). The mean PSSs were as follows: N-1 = dim: 57.91ms, N-1 = bright: 53.09ms.*

For the SJ data (see figure 6.7), the effect of preceding SOA was also apparent, with a 21ms mean PSS difference between trials preceded by a 200ms audio-leading trial, in comparison to a 200ms visual-leading trial. However, the regression lines fitted to the PSSs, dependent on the intensity of the preceding trial, appear to be closely aligned. The difference between the intercepts for trials preceded by dim (*M* = 56.93ms, *SD* = 38.26) and bright (*M* =

54.89ms, *SD* =34.43) trials was non-significant (*t* (16) = -0.42, *p* = .342, *d* = -0.10) (one-tailed).

To summarise, the results suggest that the PSS rapidly recalibrates, based on the intensity of the preceding trial. This shift was in line with our prediction in the TOJ task, though there was no effect in the SJ task.

#### 6.4.2.3. Reaction times

Figure 6.8 displays the effect of visual intensity on median RTs, across the seven SOAs. The analysis revealed a significant main effect of SOA ($F(1.49, 25.30) = 53.34$, $p < .001$, $\eta p^2 = .76$). Importantly, there was also a significant main effect of intensity ($F(1.10, 18.74) = 5.13$, $p = .033$, $\eta p^2 = .23$). Finally, the interaction between SOA and intensity was approaching significance ($F(1.13, 19.17) = 4.00$, $p = .056$, $\eta p^2 = .19$).

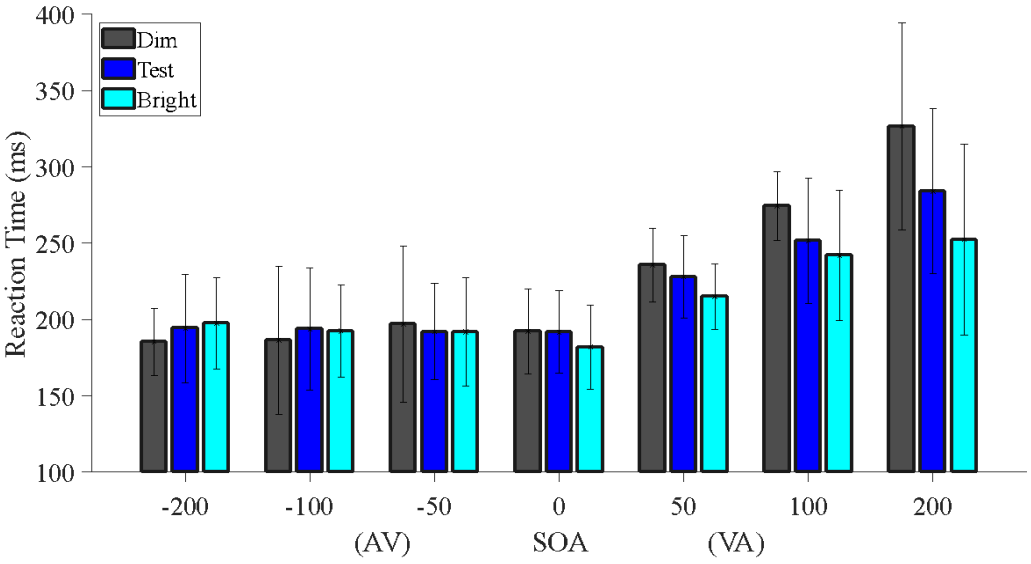


*Figure 6.8 – Bimodal reaction-times at three visual intensities (dim – grey, test intensity – blue, and bright – cyan), across all SOAs. Bars represent the average RT, calculated across each observer's median RT, for each SOA. Error bars represent standard deviations.*

The main focus of our RT analysis, however, was to assess whether the intensity or leading modality of the preceding trial affected the RT of the current trial at the test intensity (figure 6.9), indicative of a preceding-intensity and preceding-SOA-dependent effect on processing latencies. The analysis revealed a significant effect of SOA ($F$(1.09, 18.58) = 11.77, $p$ = .002, $\eta p^2$ = .41). There was, however, no effect of preceding intensity ($F$(1, 17) = 0.66, $p$ = .430, $\eta p^2$ = .04) and no effect preceding leading modality ($F$(1, 17) = 1.38, $p$ = .257, $\eta p^2$ = .08). Furthermore, no interactions were significant (all $p$ > .1).

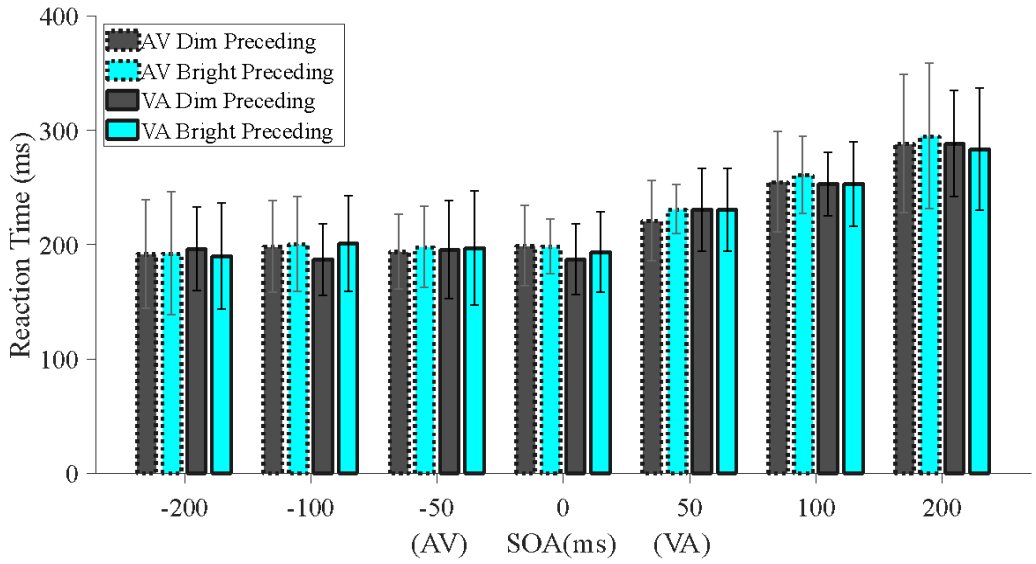


*Figure 6.9 – Bimodal reaction-times across all SOAs at the test intensity, split based on the properties of the preceding stimulus. Bars represent the average RT across each observer's median RT, dependent on whether the preceding stimulus was dim (grey bars), bright (cyan bars), auditory leading (dashed outer line) or visual leading (solid outer line). Our analysis revealed an effect of SOA, but no effect of the intensity or leading modality of the preceding trial on RTs.*

Assuming RTs are reflective of relative processing latencies, the data presented here suggest that the effect of the intensity or leading modality of the preceding trial on the PSS (in the TOJ task), is not caused by early latency shifts. This conclusion is consistent with the latency analyses on the ERP data (see section 6.5.2.5).

## 6.5. Experiment 2: ERP study

### 6.5.1. Method

#### 6.5.1.1. Participants

The ERP study involved 20 participants, aged 22-45 (*M* = 27.05, SD = 7.82). Opportunity sampling was used to obtain participants. Again, all observers reported normal or corrected-to-normal vision and hearing, and no neurological disorders.

#### 6.5.1.2. Design

A repeated measures design was used. Observers completed both the SJ and TOJ tasks on separate days, the order of which was counterbalanced.

#### 6.5.1.3. Apparatus

The apparatus were identical to experiment 1.

#### 6.5.1.4. Stimuli

The stimuli were identical to experiment 1, however only three SOAs (-150ms, 0ms, 150ms) were used, as these were deemed to provide a sufficient level of difficulty in pilot runs, while also maintaining an above chance level of correct responses. Trials were presented so that every other trial was at the test intensity, and that these trials were always preceded by a dim or bright trial (see section 6.4.1.6). Participants completed 9 separate blocks for both the SJ and TOJ task, each of which consisted of 36 trials at each of the three SOAs (324 trials per SOA in total) at the test intensity, and 18 trials at each SOA at both the dim and bright intensities (162 trials per SOA in total). There were also four filler trials per block (each consisting of three audio-visual flash-beep presentations – see section 6.4.1.6), though again, these were not included in any analyses. There were therefore 2052 trials per participant per task, and 4104 trials per participant over the course of the two sessions.

#### 6.5.1.5. Procedure

The procedure of the ERP experiment was identical to the behavioural experiment, except that participants completed only the SJ and TOJ tasks. Each task was preceded by two practice blocks, each comprising 36 trials of the highest visual intensity, split across 2 SOAs (-150ms and 150ms) in the TOJ and 3 SOAs (-150ms, 0 and 150ms) in the SJ. Participants again answered written questions after the completion of both experiments, and none reported being aware that the test intensity trials were always preceded by a trial of a different intensity.

#### 6.5.1.6. EEG data acquisition

Electroencephalographic (EEG) data was recorded using a 64-electrode BioSemi Active Two system (BioSemi, Amsterdam, Netherlands) sampled at 512Hz, with a common mode sense (CMS) used as reference. Four external electrodes were used to record electrooculograph (EOG) data – one pair was placed above and below the right eye to record vertical signals, the other was placed on the outer canthi of both eyes for the horizontal EOG signals.

#### 6.5.1.7. EEG pre-processing

The EEG data were segmented into epochs from 500ms before stimulus onset to 1000ms post-stimulus onset. The data were then bandpass filtered with lower and upper cut-off frequencies of 0.5Hz and 40Hz, respectively. The 100 ms pre-stimulus period was used for baseline correction. The SCADS procedure (Junghofer, Elbert, Tucker, & Rockstroh, 2000), widely used in the pre-processing of EEG data (e.g., Harrison & Ziessler, 2016; Johnen & Harrison, 2020; Nguyen, Kuntzelman, & Miskovic, 2017), was used to reject artefacts in the EEG data. The SCADS procedure identifies artefacts for each channel individually, then the data is transformed to average reference, and global artefacts are removed. A benefit of this pipeline is that it reduces the likelihood of artefacts from individual channels distorting

the data upon transforming it to average reference. Epochs with 10 or more contaminated channels are then rejected based on the median amplitude, standard deviation and gradient, and replaced using weighted spherical spline interpolations using data across all channels (the weighting of which depends on this distance from the reference channel to the channel which needs replacing), provided that all the rejected sensors are not within the same region (as this would result in the interpolation being invalid). Across both tasks, an average of 28.60% of epochs were identified as contaminated and removed.

#### 6.5.1.8. Statistical analysis of ERPs

To investigate the effects of preceding trial on the ERP in both tasks, the global field power (GFP) of the test intensity data was assessed separately at each SOA (figure 6.10A). In order to assess whether the effect of intensity would be evident in either early or late time-windows, and in order to limit the number of statistical tests (Luck, 2014; Luck & Gaspelin, 2017), the single highest peak in the GFP with an onset before 200ms and the single highest peak with an onset after 200ms, were selected for each SOA. These are denoted as the early time-window (ETW) and late time-window (LTW), respectively. Because of the highly similar shape of the GFP waveforms in the SJ and TOJ tasks, the same time-windows were used to analyse both tasks. The time-windows were 100-160ms (ETW) and 430-590ms (LTW) in the audio leading condition, 100-160ms (ETW) and 335-415ms (LTW) in the simultaneous condition, and 165-210ms (ETW) and 400-550ms (LTW) in the visual leading condition. The left (PO3, P1, P3, P5, CP1, CP3, CP5, C1 and C3) and right (PO4, P2, P4, P6, CP2, CP4, CP6, C2 and C4) electrode clusters used for the analysis (figure 6.10B) were chosen based on the data from Simon et al. (2017).

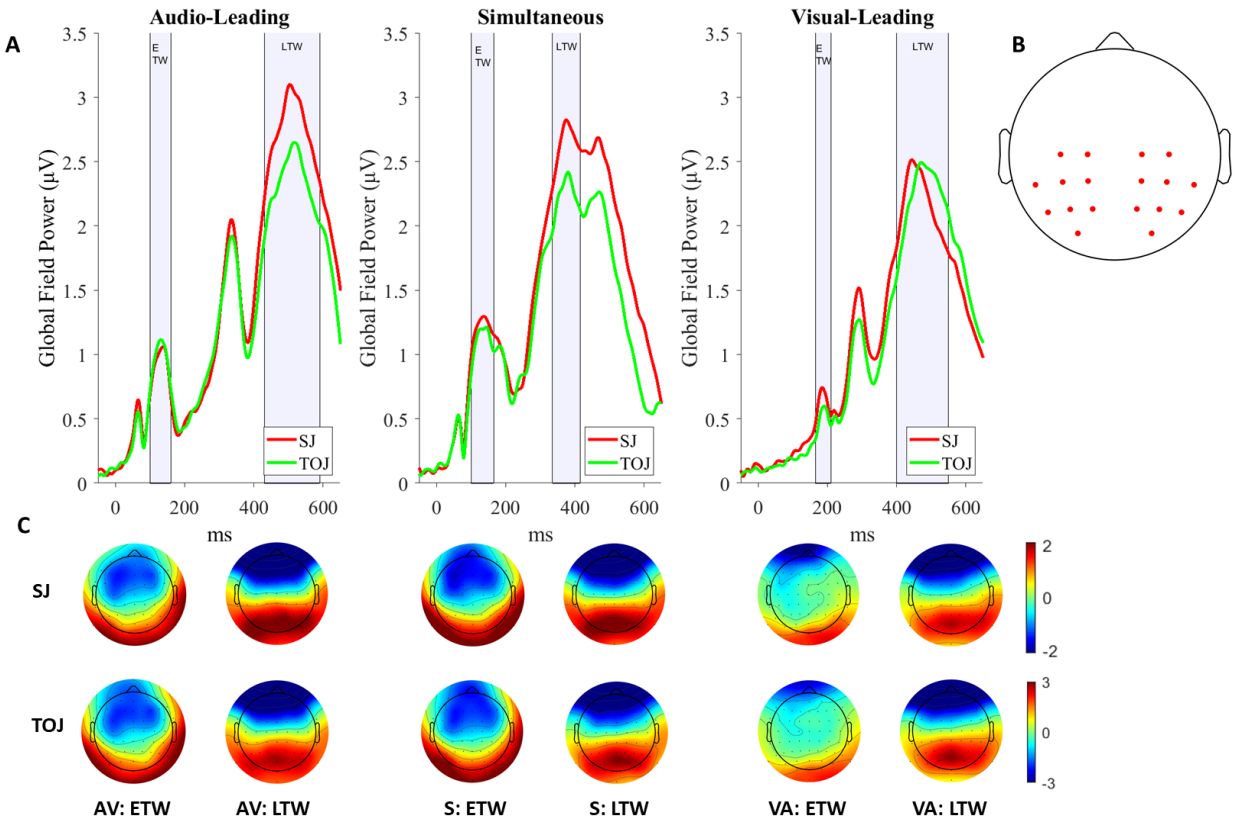


*Figure 6.10 – The global field power (A) of the test intensity data in the audio-leading (left), simultaneous (middle) and visual-leading (right) SOAs in the SJ (red) and TOJ (green) tasks. The single largest peaks which onset before and after 200ms were chosen as the early time-window (ETW) and late time-window (LTW) for the analysis. B - The two electrode clusters used in the repeated measures ANOVAs. Mean amplitudes were calculated across the left and right clusters. C – Topographic plots representing the global mean amplitude, recorded for audio-leading (AV – left column), simultaneous (S – middle column) and visual-leading (VA – right column) trials, in the SJ (top row) and TOJ (bottom row) tasks.*

In order to minimise the level of noise in the ERPs, the data recorded at the test intensity were split based on the intensity of the preceding trial only. This was done separately for the SJ and TOJ tasks. Splitting the data resulted in 6 ERP conditions for each task, one for each SOA (150ms audio-leading, 0ms, 150ms visual-leading) and preceding visual intensity (N-1 = dim and N-1 = bright). Amplitudes were averaged separately over the left and right electrode clusters, within the ETW and LTW specified for the SOA of the current trial. Repeated measures ANOVAs assessing the effect of preceding intensity (dim or bright) and electrode cluster (left or right cluster), were conducted on the mean amplitudes calculated within the ETW and LTW separately (Luck, 2014).

To assess the effect of preceding intensity on early processing latency, the ERP waveforms of test intensity trials, recorded at the Oz electrode (Bonmassar, Anami, Ives, & Belliveau, 1999; Brelén, Vince, Gérard, Veraart, & Delbeke, 2010; Thurtell et al., 2009) were analysed. The ERPs of test intensity data were split based on the intensity of the preceding trial, and the latency corresponding to the peak amplitude (e.g., Ran & Chen, 2017; Stockdale, Morrison, Kmiecik, Garbarino, & Silton, 2015) between 160-220ms was subjected to a repeated-measures *t-test*. This was done separately for the SJ and TOJ tasks, for the visual-leading trials only.

To comprehensively investigate any spatiotemporal differences between tasks, point-wise running *t-tests* were used to compare the ERPs between the SJ and TOJ tasks between 0 and 600ms. This analysis was chosen because the SJ and TOJ tasks have only been previously compared at a single electrode and at early time windows (Basharat et al., 2018). Point-wise running *t-tests* have been used extensively to investigate audio-visual processing using ERPs (Brandwein et al., 2013; Giard & Peronnet, 1999; Harrison et al., 2015; Meyer et al., 2013; Molholm et al., 2002; Stekelenburg & Vroomen, 2007). In order to reduce the prevalence of type-1 errors, clusters were only considered to be significant where 10 or more successive samples reached a significance level of $p < .05$, across 3 adjacent electrodes (Brandwein et al., 2013). Due to the notable differences in the neural processes involved in the perception of audio-leading and visual leading stimuli (Cecere et al., 2017), the audio-leading, visual leading and simultaneous trials were analysed separately.

### 6.5.2. Results

#### 6.5.2.1. Intensity dependence of the grand mean ERPs

The grand mean ERPs appear to show an intensity dependence, which is mostly consistent across the SJ (top) and TOJ (bottom) tasks; the initial early peak had a later peak latency for test intensity trials, than for bright trials, in both the SJ and TOJ tasks (figure 6.11A). A corresponding early peak was not immediately apparent for the dim trials, however.

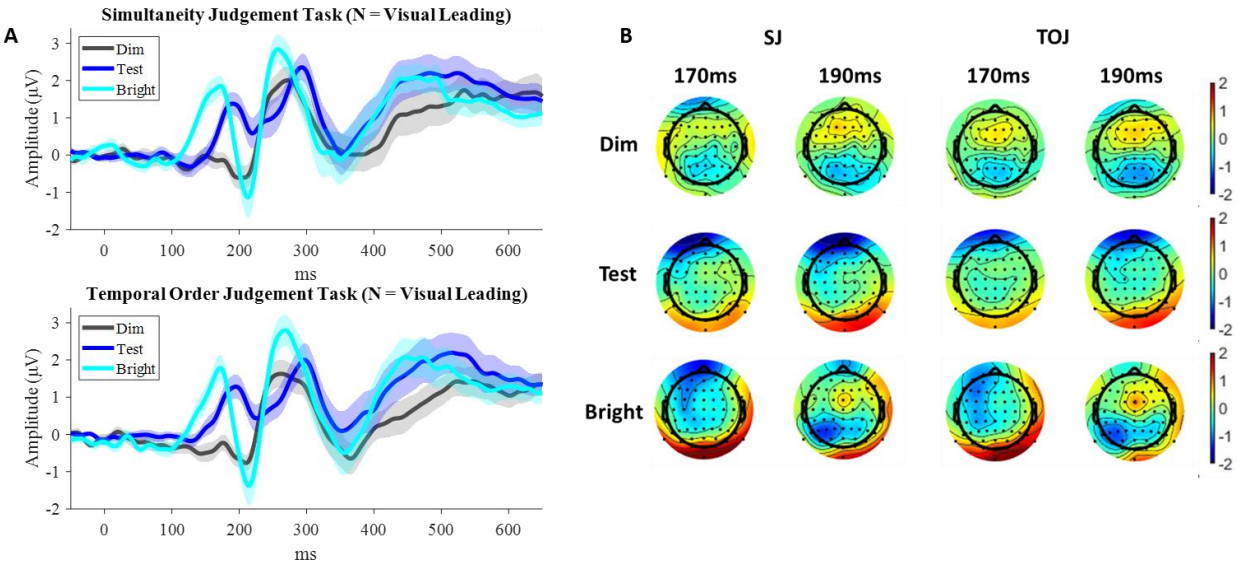


*Figure 6.11 – A) The grand-averaged ERP (+/- SEM) recorded at the Oz electrode during visual leading trials at the three visual intensities in the SJ (top) and TOJ (bottom) tasks. The earliest positive peak (~175ms) had had an earlier peak latencies (SJ = 172ms, TOJ = 172ms) and higher amplitudes (SJ = 1.85V, TOJ = 1.79V) for brighter stimuli , in comparison to the peak latency (SJ = 190ms, TOJ = 194ms) and amplitude (SJ = 1.38V, TOJ = 1.28V) for test intensity trials. No peak at this latency was immediately identifiable for the dim intensity. B) Topoplots representing the mean amplitude over all electrodes of the visual-leading condition. The topoplots represent ERP data recorded at the dim (top row), test (middle row) and bright (bottom row) visual intensities, in the SJ (columns one and two) and TOJ (columns three and four) tasks.*

#### 6.5.2.2. The effect of the preceding trial's intensity on ERP amplitude

The ERP data of the test intensity was split based on the intensity of the preceding trial (see section 6.4.2.2), in order to investigate whether the effect of preceding visual intensity was evident early or late in the ERP. Amplitudes were averaged over early and late-time-

windows, chosen using the GFP (see figure 6.10A). Figure 6.12 displays the behavioural data, both for all trials in the TOJ (A) and SJ (C) tasks, and for the test intensity trials, split based on the intensity of the preceding stimulus (B and D).

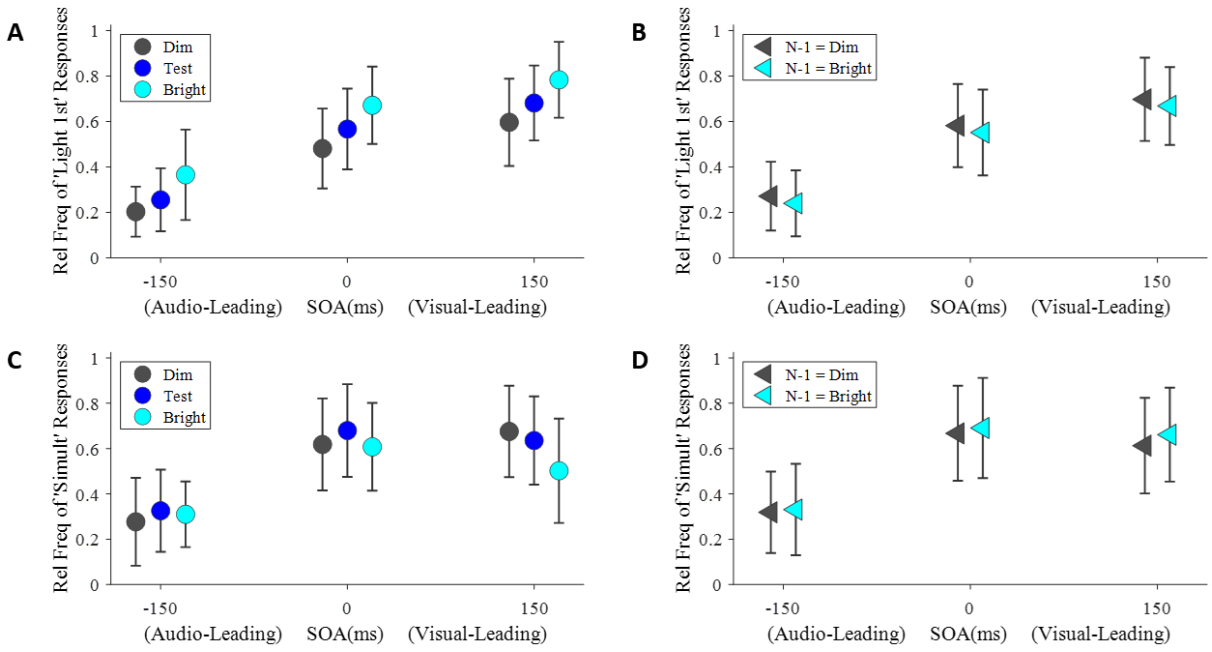


*Figure 6.12 – Behavioural data collected during the ERP experiment. The top row (A and B) displays the relative frequency of 'light first' responses (+/- standard deviation) in the TOJ task, whereas the bottom row (C and D) displays the relative frequency of 'simultaneous' responses (+/- standard deviation) in the SJ task. The left column (A and C) represents the data at the three visual intensities. The right column (B and D) represents the data collected at the test intensity, split based on the visual intensity of the preceding trial. Whilst it appears as though the intensity of the current trial affects the mean responses at each SOA (see A and C), there is seemingly only a very small effect, if any, of preceding intensity (see C and D). This may be because of the small range of SOAs used during the ERP experiment, which will likely be insensitive (at least behaviourally) to small shifts in perceived simultaneity. Note that all trials were presented at one of three SOAs (-150, 0 and 150ms); markers are offset along the x-axis for clarity.*

#### 6.5.2.3. TOJ Task

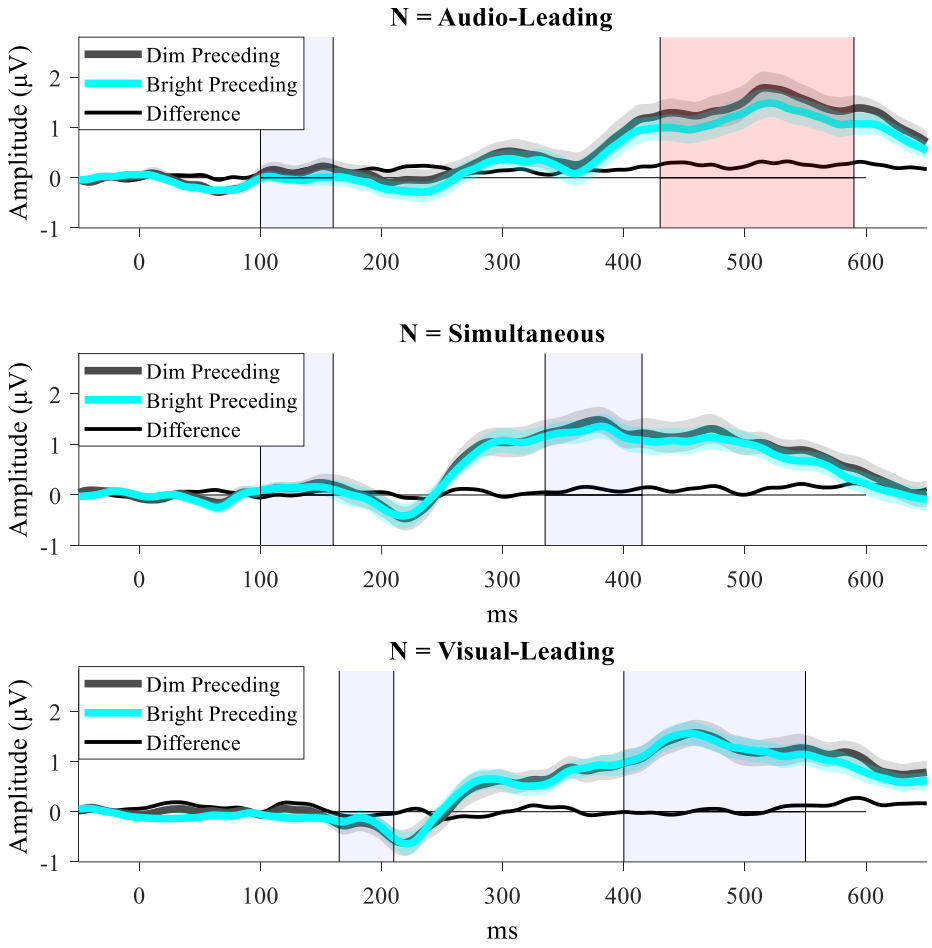


*Figure 6.13 – Grand mean ERPs (+/- SEMs) recorded in the TOJ task at the test intensity and averaged over the left electrode cluster. Each row displays the test intensity data based on the temporal structure of the current trial (N). Cyan traces represent trials preceded by a bright stimulus, whereas grey traces represent trials preceded by a dim stimulus. Black traces reflect the difference in the ERPs of the trials preceded by dim and bright stimuli. Light blue boxes highlight the early (ETW) and late time-windows (LTW) used in the analyses, that were selected based on the GFP. The red box highlights the significant effect of preceding intensity revealed by repeated measures ANOVAs.*

The grand mean ERP waveforms for the TOJ task are plotted in figure 6.13 (see also figure 6.A1). For the audio-leading SOA (figure 6.13, top row), there were no significant effects of preceding intensity or electrode cluster, and no significant interaction at the ETW (all $p >$ .05). Importantly, at the LTW, while the effect of electrode cluster on ERP amplitude was non-significant ($F(1, 19) = 0.73$, $p = .404$, $\eta p^2 = .04$), there was a significant effect of preceding intensity ($F(1, 19) = 4.93$, $p = .039$, $\eta p^2 = .21$). The interaction between preceding

intensity and electrode cluster was non-significant ($F(1, 19) = 1.72$, $p = .205$, $\eta p^2 = .08$), however.

There were no significant effects on ERP amplitude and no significant interactions, for both the simultaneous SOA (figure 6.13, middle row) and the visual-leading SOA (figure 6.13, bottom row), at either the ETW or LTW (all $p > .05$).

#### 6.5.2.4. SJ task

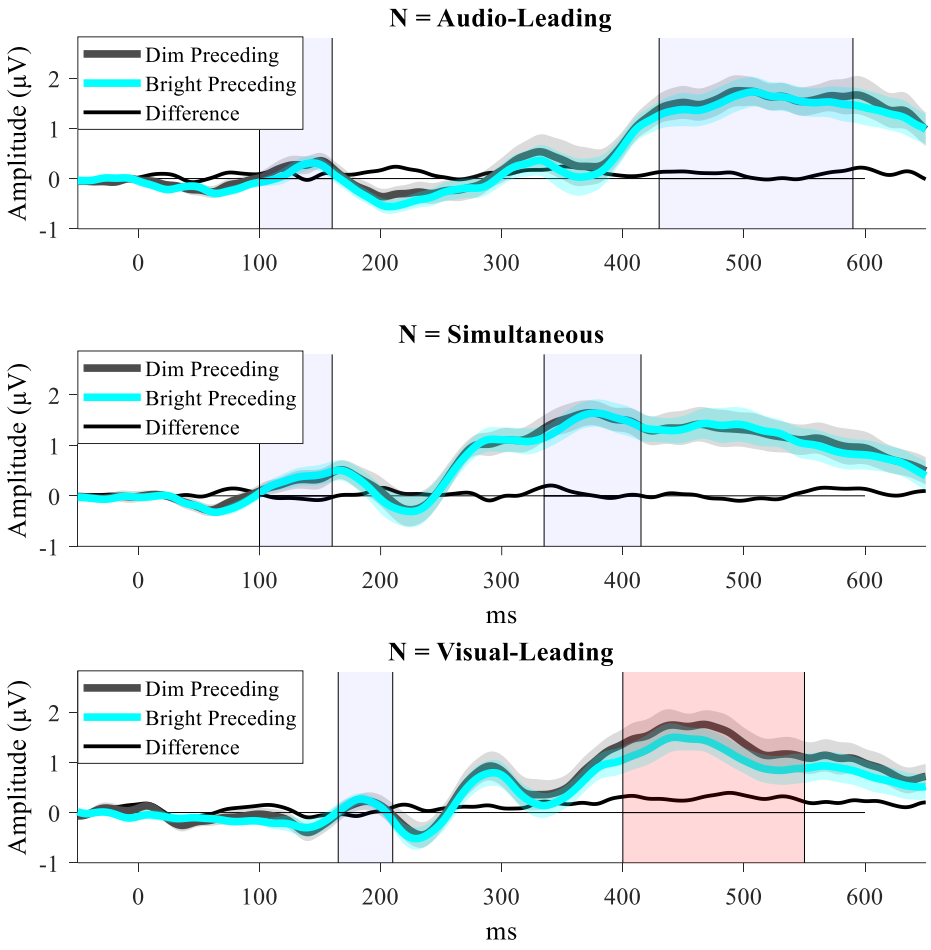


*Figure 6.14 – Grand mean ERPs (+/- SEMs) recorded in the SJ task at the test intensity, averaged over the right electrode cluster, and split based on the intensity of the preceding trial. Black traces reflect the difference in the ERPs of the trials preceded by dim and bright stimuli. Light blue boxes highlight the early (ETW) and late time-window (LTW) used in the repeated measures ANOVAs. Red boxes highlight where the significant effect of preceding intensity was evident.*

The grand mean ERP waveforms for the SJ task are plotted in figure 6.14 (see also figure 6.A2). For the audio-leading SOA (figure 6.14, top row) at the ETW, the effect of electrode cluster was significant ($F(1, 19) = 8.64$, $p = .008$, $\eta p^2 = .31$). However, the effect of preceding intensity was non-significant ($F(1, 19) = 0.64$, $p = .434$, $\eta p^2 = .03$), as was the interaction between preceding intensity and electrode cluster ($F(1, 19) = 0.66$, $p = .426$, $\eta p^2 = .03$). At the LTW, both the main effects and the interaction were non-significant (all $p > .05$).

For the simultaneous SOA (figure 6.14, middle row), there were no significant main effects on the ERP amplitude and no significant interaction, at both the ETW or LTW (all $p > .05$).

Finally, for the visual-leading SOA (figure 6.14, bottom row), there was no effect of electrode cluster or preceding intensity and no significant interaction at the ETW (all $p > .05$). However, at the LTW, there was a significant effect of electrode cluster ($F(1, 19) = 5.25$, $p = .034$, $\eta p^2 = .22$) and of preceding intensity ($F(1, 19) = 4.70$, $p = .043$, $\eta p^2 = .20$). The interaction between electrode cluster and preceding intensity was non-significant ($F(1, 19) = 3.93$, $p = .062$, $\eta p^2 = .17$), however.

#### 6.5.2.5. The effect of preceding visual intensity on ERP latency

Figure 6.15 displays the latency of the maximum amplitude of the initial positive peak (160-220ms), which was derived for visual leading trials at the Oz electrode. This was done for the test intensity ERP data, which were split based on the intensity of the preceding trial (see figure 6.11A for the grand-averaged ERP waveforms of the unsplit data). Repeated-measures *t-tests* revealed no significant effect of preceding intensity on processing latency in both the SJ ($t(19) = -0.62$, $p = .542$, $d = 0.17$) and TOJ ($t(19) = -1.17$, $p = .256$, $d = 0.31$) tasks.

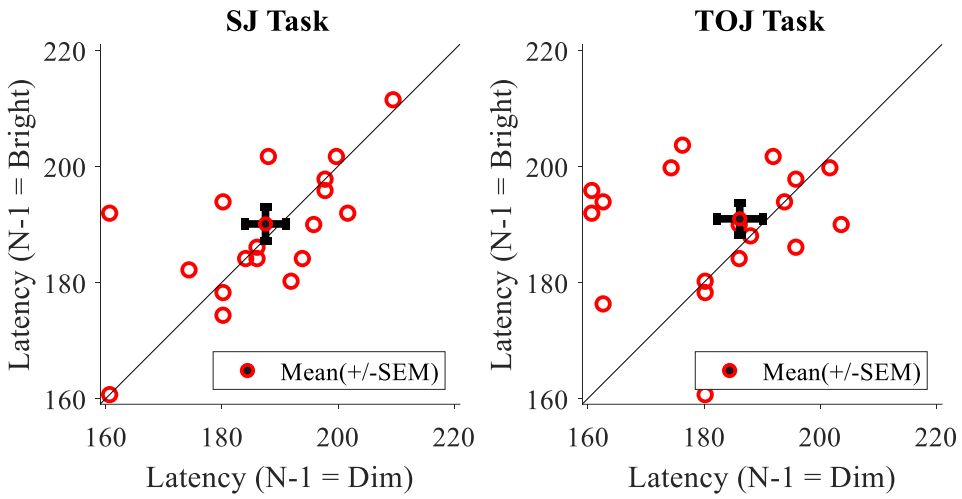


*Figure 6.15 – Scatterplots of the latency of the initial positive peak of the ERP in the SJ (left) and TOJ (right) task. Data represents the visual leading condition of the test intensity, recorded at the Oz electrode. Data points to the right of the line indicate a longer latency when the preceding trial was dim.*

#### 6.5.2.6. Differences in the ERP of the SJ and TOJ tasks

It is thus far unclear whether the lack of correlation between the PSSs reported in our behavioural experiment (see section 6.4.2.1) reflects differences in the processing mechanisms underlying the tasks (Love et al., 2013; Zampini et al., 2003) or simply differences at the decisional stage (Parise & Ernst, 2016). This section aims to resolve this uncertainty using ERPs.

Exploratory, point-wise running t-tests were used to compare the two tasks at each SOA, in order to holistically analyse the spatiotemporal differences underlying the tasks. The ERPs recorded at the test intensity for the two tasks are displayed in figure 6.16. It is clear that the neural signatures of the two tasks match very closely across the early timepoints, and

only differ after 400ms.

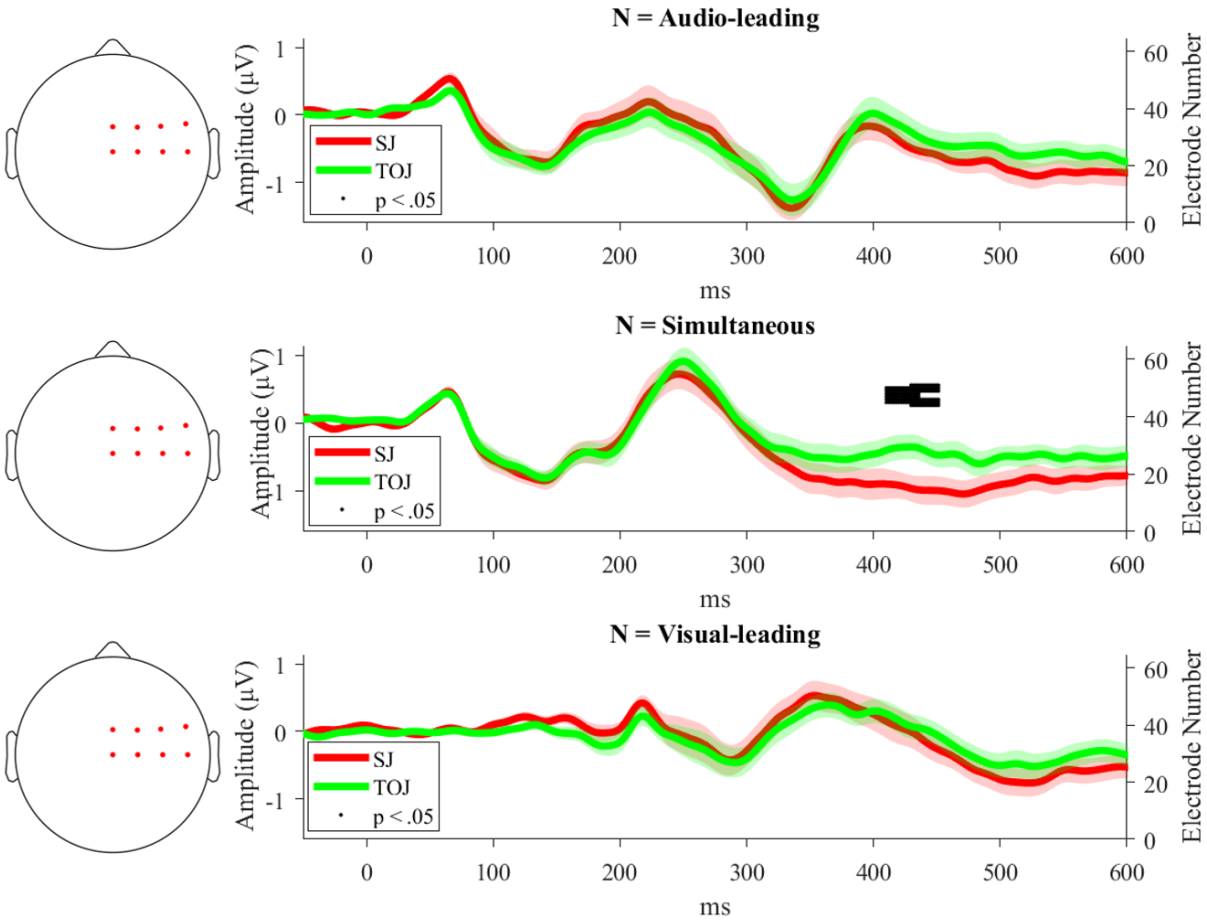


*Figure 6.16 – The grand average ERP of the SJ (red) and TOJ (green) task across the audio-leading (top), simultaneous (middle) and visual-leading (bottom) SOAs. Faded regions represent +/- standard error of the mean. The significant time points highlighted using point-wise running t-tests are represented by black dots. The ERP waveforms (left column) were averaged across the electrodes highlighted in the left column, which were chosen based on the significant electrodes in the t-tests at the simultaneous SOA.*

The analysis revealed a cluster of significant electrodes at late time points for the simultaneous SOA. This comprised significant time points across a range of central and fronto-central electrodes between 410-451ms. There were no differences for the audio-leading and visual-leading SOAs, however. 'Late' differences in the ERP are typically ascribed to decisional processes (e.g., Kelly & O'Connell, 2013), rather than differences in early sensory processing stages (e.g., Basharat et al., 2018). The data presented here comply with the former.

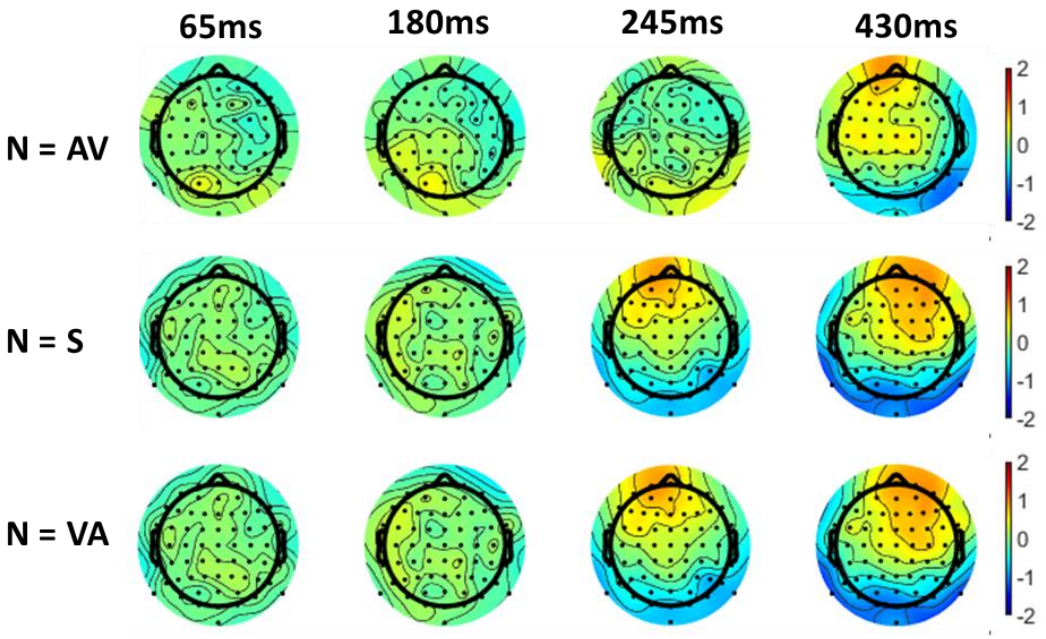


*Figure 6.17 – Topoplots representing the difference in the grand mean ERP between the SJ and TOJ task (TOJ - SJ), for audio-leading (N = AV), simultaneous (N = S) and visual-leading (N = VA) trials at four time points of interest, selected based on the peaks in figure 6.10A.*

### 6.6. Discussion

The current paper aimed to investigate the hypothesis that the perception of simultaneity rapidly recalibrates to compensate for relative, intensity-dependent central arrival times of audio-visual signals. It was predicted that dim stimuli would be processed at longer latencies than bright stimuli, and therefore the perception of simultaneity would shift to compensate for the resultant difference in central arrival times. This was predicted to result in a PSS shift towards the visual-leading SOAs following bright stimuli, and towards auditory-leading following dim stimuli. Experiment 1 partially supported this hypothesis as the PSS rapidly recalibrated based on the intensity of the preceding trial, and the direction of this shift was in line with our prediction. However, this was only evident in the TOJ task. Experiment 2 aimed to investigate whether this putative effect was driven by early processing differences, or differences at the decisional stage, by assessing the effect of the visual intensity of the preceding stimulus on the ERP of the current trial. This revealed a

significant effect of preceding intensity in both the SJ and TOJ tasks. Importantly, these effects were only evident late in the ERP (>300ms) across central and parietal regions. Additionally, there was no effect of preceding intensity on ERP latency, and no effect of either preceding intensity or preceding leading modality on the simple RTs reported in experiment 1, consistent with the prediction that rapid recalibration is not driven by changes in early processing latencies.

A secondary aim of this paper was to assess the differences in the time course of the neurocognitive processes underlying the SJ and TOJ tasks. The comparison of the SJ and TOJ tasks in experiment 2 revealed only late differences in the ERP (> 400ms), consistent with the prediction that the tasks differ only at the decisional stage (e.g., Parise & Ernst, 2016), rather than during early sensory processes.

### 6.6.1. The important role of intensity-dependent processing latencies in the perception of simultaneity

Processing latencies, estimated using both response times (Pöppel et al., 1990) and ERPs (Plainis et al., 2013) are dependent on visual intensity. Our reaction time data showed a similar intensity dependence (see figure 6.8), most notably for the visual-leading trials, suggesting observers responded after detecting the first arriving signal, at least at the larger SOAs (Raab, 1962). Likewise, the PSSs estimated using the SJ and TOJ tasks shifted towards the visual leading SOAs with reducing visual intensity, consistent with previously reported data (Boenke, Deliano, & Ohl, 2009; Horsfall et al., 2021; Jaśkowski, 1992; Leone & Mccourt, 2015; Menendez & Lit, 1983; Roufs, 1963). Taken together with the intensity-dependence evident in the ERP recorded over the visual cortex (see figure 6.11), these findings contribute to the argument that the intensity dependency of perceived simultaneity is driven by relative intensity-dependent arrival latencies at multisensory brain regions (Leone & McCourt, 2015; Horsfall et al., 2021).

### 6.6.2. The inconsistent effect of preceding intensity on perceived simultaneity

The PSS has been shown to rapidly recalibrate due to the leading modality of an immediately preceding trial (Noel et al., 2016; Roseboom, 2019; Simon, Noel, Wallace, et al., 2017; Van der Burg et al., 2013; Van Der Burg et al., 2015). This effect was also evident in our data; following a visual-leading stimulus, the PSS shifted towards the visual-leading SOAs, and the opposite was true following audio-leading stimuli. Furthermore, the magnitude of this shift was greater for trials preceded by larger offsets.

While the effect of preceding SOA on the PSS is widely replicated, the reason for such a mechanism to exist is not known. Here we propose that a central purpose of rapid recalibration is to correct for rapidly varying stimulus intensities, in this case luminance. Since the perception of simultaneity shifts based on the physical offset of a stimulus, it was hypothesised that intensity-dependent latency delays would induce an equivalent shift in the PSS (see figure 6.1).

For the TOJ data, a preceding-intensity-dependent PSS shift was observed, and critically, the direction of this shift was in line with our prediction. Van der Burg et al. (2018) investigated how rapid recalibration (based on preceding SOA) relates to the perceived (i.e., reported) temporal order of the stimuli, by asking observers to perform TOJs and SJs on alternate trials. Their results suggested that rapid recalibration is based solely on physical audio-visual offsets, rather than perceived offsets. Our behavioural findings are inconsistent with this, by showing that rather than being dependent simply on the physical temporal offsets of stimuli, rapid recalibration is dependent on relative (intensity-dependent) *physiological* arrival latencies.

In contrast to the TOJ data, no effect of preceding intensity was observed on the PSS when estimated using the SJ task. Such a disparity with the TOJ data is not necessarily surprising, given the previously reported differences in task difficulty (Love et al., 2013), decisional

criteria (Yarrow et al., 2011) and response biases (Linares & Holcombe, 2014), as well as the inconsistency in the effect of preceding SOA on the PSS when estimated using the two tasks (e.g., Recio et al., 2019; Roseboom, 2019). The current data therefore add to the vast array of differences highlighted between the tasks. While the reason for the inconsistency between the current data of the two tasks is uncertain, it has been previously argued that it is possible to detect more subtle changes in the PSS using the TOJ task, in comparison to the SJ task, due to the higher cognitive demand required to make TOJs (Shore, Gray, Spry, & Spence, 2005). It therefore may be possible that the current SJ design is less sensitive to the small PSS shifts that are evident in the TOJ data.

### 6.6.3. Rapid recalibration to preceding intensity is consistent with the decisional process underlying rapid recalibration to preceding SOA.

The significant difference in the ERPs of trials preceded by bright and dim stimuli in the TOJ task, highlighted in figure 6.13 (top row), maps onto the audio-leading side of the psychometric functions fitted to the behavioural data in figure 6.6B. Likewise, the difference in the ERPs of the SJ data (figure 6.14, bottom row), maps onto the visual-leading side of the psychometric functions in figure 6.7B. If the behavioural and ERP data were entirely consistent, we would perhaps expect an effect of preceding intensity only on the audio-leading data points (i.e., the relative frequencies plotted for each SOA) in the TOJ task (figure 6.6B) and the visual-leading data points in the SJ task (figure 6.7B), though this was not apparent.

In the TOJ task, an effect of preceding intensity was evident only late in the ERP (> 300ms) of audio-leading trials, and was observed across parietal, centro-parietal and central electrodes. Importantly, the effect of preceding intensity highly corresponds with that reported by Simon et al. (2017), who found a significant effect of preceding leading modality (audio-leading or visual-leading) on the ERP, across overlapping time points and

electrodes. Simon et al. (2017) concluded that the preceding-SOA-dependent shifts were driven by decisional processes, as their findings closely aligned with previously described late ERP components involved in perceptual decision making (Kelly & O'Connell, 2013; O'Connell et al., 2012; Tagliabue et al., 2019). The findings of the current EEG experiment suggest that preceding-*intensity*-dependent rapid recalibration is driven by a consistent decisional mechanism.

In the case of the SJ task, preceding-intensity-dependent ERP differences were observed for visual-leading trials. The disparity between the behavioural and ERP data of the SJ task may be driven by the more sensitive nature of ERP measures. It is possible that the high temporal acuity of EEG identified a preceding-intensity-dependent effect that could not be detected behaviourally with the current design. An alternative explanation for the observed differences in the ERPs is that parieto-occipital regions have been previously shown to have a top-down role in the modulation of perceived visual brightness (Zhou et al., 2020), perhaps through the involvement of feedback mechanisms into V4 (Roe et al., 2012). It is possible that the differences, here attributed to rapid recalibration, were instead due to the process described by Zhou et al. (2020). Future research using a neuroimaging technique with a higher spatial resolution, such as magnetoencephalography, should be conducted to contribute to this discussion further.

#### 6.6.4. Rapid recalibration to preceding visual intensity is not driven by early latency shifts

If rapid recalibration was a low-level adaptation effect, we would expect to see an effect of the preceding trial's characteristics on early processing latencies and reaction times. While we observed a clear effect of intensity in our simple RT task, there was no effect of preceding leading modality or intensity on RTs. Consistent with our RT data, additional analyses of our ERP data revealed no effect of preceding intensity on early processing

latency, in either the SJ or TOJ task. As far as we are aware, ours is the first experiment to investigate the effect of rapid recalibration (i.e. a single preceding trial) on RTs, and our findings are at odds with the possibility that the properties of a single preceding trial (which affect relative central arrival times), are compensated for via a change in relative processing latencies, such as what is observed following prolonged exposure to temporally offset bimodal stimuli (Di Luca et al., 2009; Navarra et al., 2009).

### 6.6.5. Differences in the neural time courses underlying the SJ and TOJ tasks

The PSSs estimated from the SJ and TOJ tasks are commonly found to be uncorrelated (Horsfall et al., 2021; Love et al., 2013; Van Eijk et al., 2008), an effect which was also evident in our data at all three visual intensities. It is currently disputed whether this discrepancy is due to different underlying processing mechanisms (Jaśkowski, 1991; Love et al., 2013; Zampini et al., 2003), or differences in the way identical early signals are utilised at the decisional stage (Parise & Ernst, 2016). Our data highlight late differences in the ERP of the two tasks (>400ms) across central and fronto-central electrodes. Corresponding electrodes and time-windows have been associated with performance monitoring and task difficulty in perceptual decision making (Pardo-Vazquez, Padrón, Fernández-Rey, & Acuña, 2014). Our data are inconsistent with previous ERP research, which showed only early differences in the ERP in an audio-leading condition (Basharat et al., 2018), though it must be noted that the authors did not report any analyses of the late ERP. The two tasks are highly dependent on stimulus parameters (e.g., Kanai, Dalmaijer, Sherman, Kawakita, & Paffen, 2017; Zampini et al., 2005, 2003) and differences between observers (e.g., Basharat et al., 2018; Love et al., 2018), which may have contributed to the difference between the current data and Basharat et al.'s (2018) findings. Crucially, the task-dependent differences reported in our data were isolated to the late ERP, consistent with the proposal that the two tasks vary only at the decisional stage (Parise & Ernst, 2016).

### 6.7. Chapter summary

The current experiments highlight that the PSS, when estimated using the TOJ task, rapidly recalibrates based on the visual intensity of the preceding trial. While this effect was small (< 20ms), it was larger than the predicted preceding-intensity-dependent shift. Experiment 2 revealed a late (> 400ms) effect of preceding intensity on the ERP of both the TOJ and SJ tasks, suggesting that this compensatory process occurs at the decisional stage, rather than being driven by a low-level adaptation effect. However, statistically significant ERP differences were only evident at one of three SOAs for both the SJ and TOJ tasks, with relatively small effect sizes. The lack of convergence between the behavioural and ERP data raises questions on the robustness of the data, since an effect of preceding intensity was not evident in experiment 1 for the SJ task, and since the differences observed in experiment 2 do not closely map onto the behavioural data reported in experiment 1 (see section 6.6.3). It is therefore clear that the data should be interpreted with some caution.

Taken together with Simon et al.'s (2017) findings, the current chapter provides some evidence to suggest that the PSS rapidly recalibrates to visual intensity-dependent processing latencies and physical temporal offsets through a consistent decisional process. Since other cues, such as stimulus size and distance, can be used in perceptual judgements, further research using more ecologically valid stimuli will be required to see if such an effect occurs in real-world settings.

The plasticity of the PSS in response to an immediately preceding trial was evident in the current chapter, as was the effect of intensity on perceived simultaneity. The following chapter discusses the role of prolonged training on TBW plasticity and the generalisability of training-induced changes across stimuli with different visual intensities.

## 6.8. Appendix 6

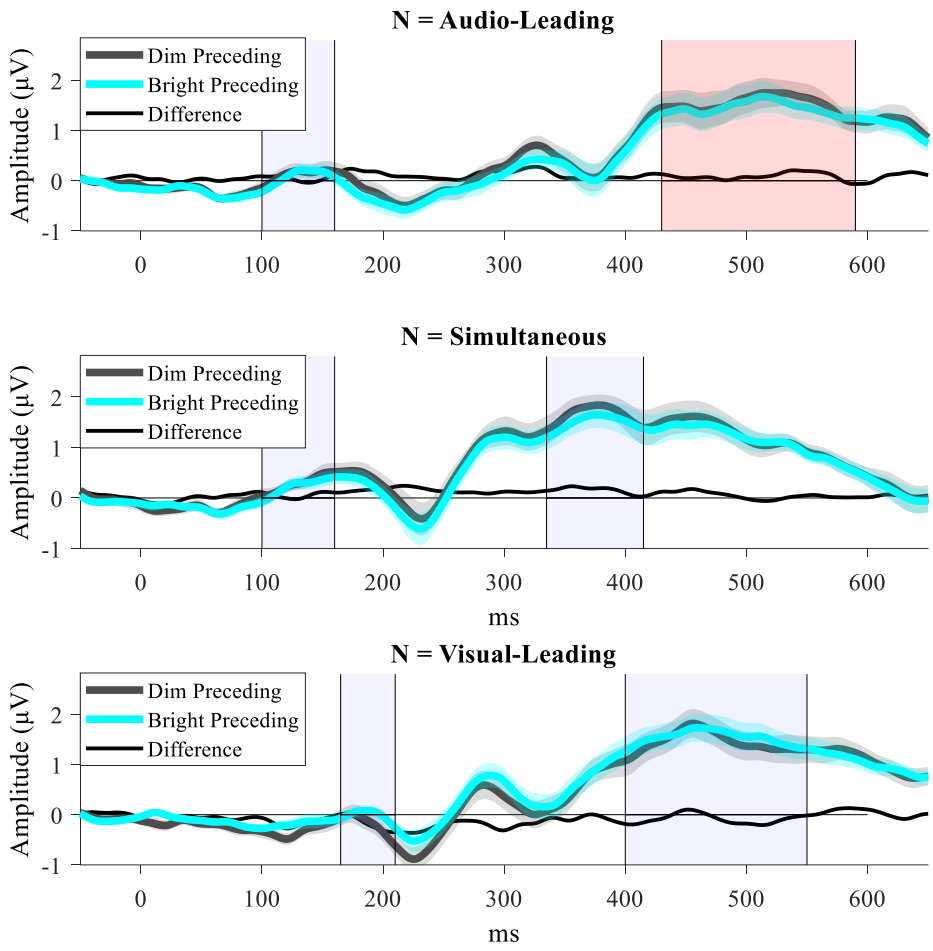


*Figure 6.A1 – Grand mean ERP recorded over the right electrodes, in the TOJ task at the test intensity. Waveforms represent the data split based on the intensity of the preceding trial.*

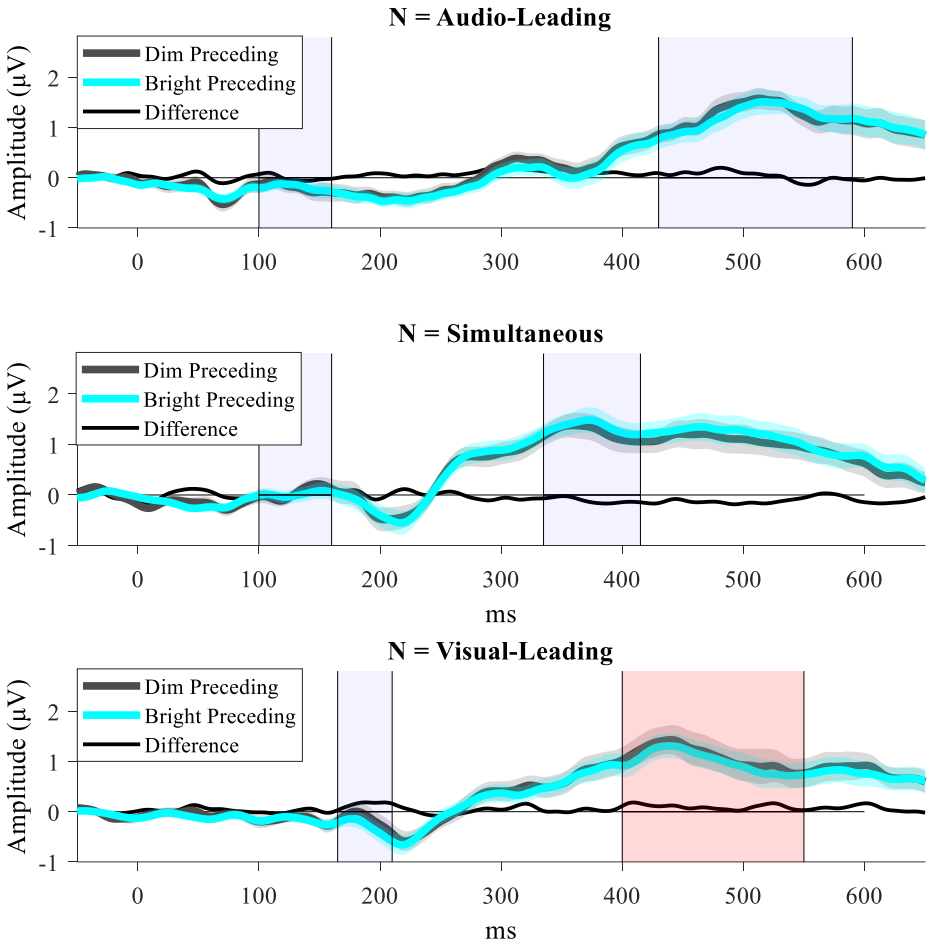


*Figure 6.A2 – Grand mean ERP recorded in the SJ task at the test intensity, recorded over the left electrodes, split based on intensity of the preceding trial.*

# 7.Narrowing of the audio-visual temporal binding window due to perceptual training is specific to high visual intensity stimuli.

## 7.1. Chapter introduction

The results of chapter 4, chapter 5 and chapter 6 have highlighted the effect of visual intensity on perceived simultaneity. Furthermore, chapter 6 highlighted the plasticity of the PSS, based on the intensity and SOA of the preceding trial.

Previous research has highlighted the plasticity of the TBW in response to training in the SJ task (Powers et al., 2009). SJ training has therefore been proposed as a potential intervention to reduce the size of the TBW in those with multisensory perceptual deficiencies (Foss-Feig et al., 2010; Powers et al., 2009; Setti et al., 2014; Stevenson et al., 2017; Wallace & Stevenson, 2014). However, the generalisability of perceptual learning-related performance improvements across stimuli with different properties is inconsistent (e.g., Fiorentini & Berardi, 1980; Green, Kattner, Siegel, Kersten, & Schrater, 2015; McGovern, Roach, & Webb, 2012). The current chapter therefore investigates the generalisability of training-induced reductions in TBW size across stimuli with different visual intensities.

### 7.2. Abstract

The temporal binding window (TBW), which reflects the range of temporal offsets in which audio-visual stimuli are combined to form a singular percept, can be reduced through training. Our research aimed to investigate whether training-induced reductions in TBW size transfer across stimulus intensities.

32 observers performed simultaneity judgements at two visual intensities with a fixed auditory intensity, before and after receiving audio-visual TBW training at just one of these two intensities.

We show that training individuals with a high visual intensity reduces the size of the TBW for *bright* stimuli, but this improvement did not transfer to dim stimuli. The reduction in TBW can be explained by shifts in decision criteria. Those trained with the dim visual stimuli, however, showed no reduction in TBW.

Our main finding that perceptual improvements following training are specific for high-intensity stimuli, potentially highlighting limitations of proposed TBW training procedures.

### 7.3. Introduction

Perceptual learning refers to improvements in the processing of sensory information brought on through experience. Behavioural performance improvements (Fahle, 1997; Herzog & Fahle, 1997) are typically complemented by structural and functional changes in primary sensory cortices (for a review see: Fahle & Poggio, 2002). However, perceptual learning has been shown to affect both perceptual sensitivity (McGovern et al., 2012) and decision criteria (Bang & Rahnev, 2017; Herzog, Ewald, Hermens, & Fahle, 2006).

Recent research on the integration of auditory and visual stimuli suggests that multisensorial binding mechanisms are plastic and can be modulated by rapid (Noel, De Niear, Van der Burg, & Wallace, 2016; Simon, Noel, & Wallace, 2017; Van der Burg, Alais, & Cass, 2013) and sustained exposure (Di Luca et al., 2009; Fujisaki et al., 2004; Vroomen et al., 2004). Temporal proximity between the auditory and visual stimuli dictates whether these stimuli are integrated into a unified percept, i.e. the closer in time the stimuli are, the more likely it is that they are combined. This temporal binding window (TBW) within which audio-visual stimuli are perceived as simultaneous is obtained with a simultaneity judgement (SJ) task where observers indicate whether auditory and visual stimuli with varying temporal offsets appear simultaneous.

From a functional perspective it is desirable to discard low-level stimulus features such as stimulus intensity when judging whether an auditory and a visual event are simultaneous, since these stimulus attributes affect peripheral processing, but do not reflect physical simultaneity. Light travels considerably faster than sound, so that two simultaneously emitted stimuli will be detected by sensory cells with a distance-dependent time delay. Furthermore, the transduction between sensory cells and primary cortices also varies by modality, with considerably shorter latencies in the auditory than in the visual system (Raij et al., 2010). In addition to signal reception delays, there are differences in processing

times as a function signal frequency in the auditory domain (Woods, Alain, Covarrubias, & Zaidel, 1993) or colour for visual signals (Pollack, 1968). A major determinant for processing delays is signal intensity (Lakhani, Vette, Mansfield, Miyasike-daSilva, & McIlroy, 2012; Nissen, 1977), evident in the intensity-dependency of the point of subjective simultaneity (L. M. Leone & Mccourt, 2015), which reflects the stimulus offset asynchrony (SOA) at which audio-visual stimuli are most likely to be perceived as simultaneous.

The TBW absorbs some of the variability in arrival times brought about by peripheral processing differences (for a review see: Wallace & Stevenson, 2014). The TBW is plastic across the lifespan (Hillock-Dunn & Wallace, 2012; Noel et al., 2016) and depends on stimulus modality (Noel, Wallace, Orchard-Mills, Alais, & Van Der Burg, 2015) and complexity (Stevenson et al., 2014). Interestingly, individuals with schizophrenia (Stevenson et al., 2017) and autism spectrum disorders (Foss-Feig et al., 2010; Stevenson et al., 2014) have been shown to have significantly wider TBWs than controls, leading to impairments in speech perception (Stevenson et al., 2018).

Using feedback, Powers, Hillock and Wallace (2009), demonstrated a significant reduction in TBW size after a single training phase, with a mean reduction of nearly 80 ms. Whether these training effects are brought about by an improvement in temporal processing in the auditory (De Niear et al., 2016) or the visual modality (Powers et al., 2009), or both, is still an open question. Furthermore, the degree to which audio-visual simultaneity training is task and stimulus specific, is still unclear. TBW training using the SJ task does not transfer to improvements in a comparable perceptual task with identical stimuli (Matthews et al., 2016), and does not reduce susceptibility to the sound-induced flash illusion, an alternative measure of audio-visual temporal acuity (Powers et al., 2016). However, training at a temporal order judgement task, with unimodal stimuli, has been shown to cause reductions in bimodal TBW size measured using the SJ task (Stevenson et al., 2013).

A hallmark of perceptual learning is stimulus specificity. The aim of our research was to investigate whether training-induced reductions in TBW size transfer across stimulus intensities. Our hypothesis was that training will reduce the TBW size, but this learning effect will not transfer across intensities.

To test our hypothesis, participants were trained on either a low or high visual intensity stimulus paired with an auditory stimulus of fixed intensity. Participants were tested with both stimulus intensities before and after training. Our main finding is that TBW size is reduced with training, but this training effect is specific to high visual intensity stimuli.

## 7.4. Method

### 7.4.1. Participants

32 participants (23 female), aged 18-28 (mean = 21.03, SD = 2.44), were recruited via opportunity sampling. All reported normal or corrected to normal vison and normal hearing. 2 participants were removed from the final analysis as their relative frequency of 'simultaneous' responses did not exceed 0.1 at any SOA in either the pre-training or post-training dim visual stimuli condition. Participants had to achieve the following minimum performance requirements after training to be included in the analysis: 1) an overall TBW of below 1000 ms, and 2) a peak in the fitted curve between -300 and 300 ms, the range of tested values. Of the remaining participants, 11 completed the bright training condition and 10 were trained with the low-intensity stimuli.

### 7.4.2. Design

Participants were assigned to either the bright or the dim visual stimuli training group, and performed a simultaneity judgment task consisting of interleaved bright and dim visual stimuli both before and after the assigned training condition. The stimuli in the pre- and post-training sessions were identical in both groups.

### 7.4.3. Apparatus

Participants were seated in anechoic chamber (IAC, Winchester, UK), 113 cm from a LED and speaker that were held at roughly eye level in an adjustable clamp. A Tucker Davies RP2.1 real-time processor (TDT technologies, Alachua, FL) was used to generate the visual and auditory stimuli. A single 'Xenta M-219 Notebook speaker' produced the auditory stimuli. This was located 1.62° below a 5 mm white LED that was attenuated using three neutral density filters with a fractional transmittance of 50%, 25% and 6.25% respectively. A custom-built button box was used to record the participant's responses. MATLAB (ver. R2017b) was used on a PC located outside of the booth to control the Tucker Davies system and record responses.

### 7.4.4. Threshold Estimation

Using identical apparatus to the current experiment, 8 participants (age range: 22-43, mean = 27.00, SD = 2.39) completed a visual threshold estimation, followed by an auditory threshold estimation using a two-interval forced choice procedure, after 15 minutes dark adaption. Two interleaved staircases with the QUEST (Watson & Pelli, 1983) threshold estimation procedures were used for each modality. In the visual threshold estimation, participants were presented with two clearly audible 100ms beeps, with an SOA of 1 second. A 100ms visual flash occurred 250ms after one of the two beeps (Koenig & Hofer, 2011). The auditory threshold estimation mirrored that of the visual threshold, in that an auditory beep followed one of two clearly visible flashes. In each task, participants were asked to press the left or right button, depending on whether the target stimulus followed the first or second non-target stimulus, respectively. The final threshold was estimated to be the minimum intensity at which the participant could detect the stimulus with ~75% accuracy.

### 7.4.5. Stimuli

The stimuli used in the main experiments consisted of a 100 ms visual flash and a 100 ms auditory beep. The visual flash was presented at two intensities (0.02 cd/$m^2$ and 1.34 cd/$m^2$), which were 6 db and 24 db above the mean estimated visual threshold (see threshold estimation). The auditory beep had a fixed frequency of 1000 hz and was presented with a flat tone amplitude envelope at 35.54 db, 15 db above the mean estimated threshold. During the training sessions, auditory beeps were used to provide feedback after each response from the participant. A correct response was defined as saying 'simultaneous' for an SOA of 0 ms. A correct response was followed by two 1500 hz tones, lasting 40 ms each and separated by 100 ms. An incorrect response was signified by a single 150 ms tone at 500 hz. The inter-stimulus interval of all trials was 1500 ms, plus a random value between 0-2000 ms.

### 7.4.6. Simultaneity Judgement Task

Bimodal trials were presented at 11 stimulus onset asynchronies (SOAs) (-300, -200, -150, -100, -50, 0, 50, 100, 150, 200 and 300 ms), with positive values indicating that the visual stimulus was leading. The stimuli were randomly interleaved and were presented 20 times at both intensity levels (bright and dim) and at each SOA, both before *and* after training. These trials were split into two pre-training and two post-training blocks, totalling 880 trials for each participant.

### 7.4.7. Training Task

Bimodal trials were presented at either the bright or dim visual intensity, and at only 7 SOAs (-150, -100, -50, 0, 50, 100 and 150 ms). Training induced reductions in TBW size are dependent on the trained SOAs being suitably small (De Niear et al., 2016), and previous research has shown that a single training session at these offsets was enough to reduce

TBW size (Powers et al., 2009). There were 600 trials presented in a random order across two training blocks. Within each block of the training phase the ratio of simultaneous to non-simultaneous trials was 1:1, where the non-simultaneous trials were split evenly across the 6 non-simultaneous SOAs. The 1:1 ratio aimed to reduce the likelihood that the training would instil a bias, causing participants to simply reduce their overall 'simultaneous' response rate, and inadvertently reducing their estimated TBW size (Powers et al., 2009). Feedback was given after every response by the participant (see section 7.4.5). The auditory feedback was presented after every response, within 150 ms of this response. No feedback was given if participants failed to respond within 2750 ms of a stimulus offset.

### 7.4.8. Data Analysis

The data were fitted using Yarrow, Jahn, Durant and Arnold's (2011) procedure. The bell-shaped simultaneity data were fitted using the difference of two cumulative Gaussians, allowing for an asymmetrical fit for the visual leading (VA) and audio-leading (AV) responses. For each side (AV, VA) two parameters are fitted, the mean and the standard deviation of the cumulative Gaussian. Following Yarrow et al. (2011), the standard deviation is used as a measure of sensory noise (plus any variance in criterion placement). The TBW is defined as the difference between the two means (criteria) of the gaussian curves.

2 x 2 ANOVAs were used to assess the effect of intensity (bright or dim) and training (pre- or post-training) on temporal binding window size (ms). Additionally, 2 x 2 x 2 ANOVAs were used to assess the effect of intensity, training and leading modality (audio-leading or visual leading) on the SDs of the fitted Gaussians. Additional post-hoc comparisons were carried out using Wilcoxon Signed-ranks tests.

### 7.4.9. Procedure

The experiment began with two practice blocks consisting of 36 bright trials, split across three SOAs (-200 ms, 0 ms and 200 ms), in which participants were instructed to press the top button if they believed that the two stimuli occurred simultaneously and the bottom button if they occurred non-simultaneously. Subsequently, participants were given a 15-minute dark adaption period, followed by two SJ blocks, and a single training block. Participants were then given a 15 minutes break, during which they could leave the anechoic chamber. This break aimed to reduce the impact of boredom/fatigue on post-training performance. Participants then completed an additional 15-minute dark adaption, followed by the second training block, and two post-training SJ blocks.

## 7.5. Results

The primary purpose of this study was to establish whether any training-induced reductions in TBW size transfer across stimulus intensities. To test this hypothesis we measured the TBW before and after training with either a bright (high-intensity) visual stimulus (Figure 7.1) or a dim (low-intensity) visual stimulus (Figure 7.2), and a constant auditory intensity. Both groups were then tested with dim and bright stimuli.

Each individual observer's data were fitted with a difference of two cumulative Gaussians (Yarrow et al., 2011). The relative frequency of 'Simultaneous' responses is plotted as a function of SOAs, with audio-leading (A-V) on the left and visual leading (V-A) on the right side. The first row shows the data and best fitting curves pre-training, the 2$^{nd}$ row for post-training. There is substantial variability between observers which is also reflected in the estimated TBW and the criteria (row 3). If there is no effect of training, the TBWs and criteria should be clustered around the 45-deg line; if training takes place the TBW data points should lie below this line. For the criteria, a reduction in TBW size should be

reflected in the AV criteria to lie above the 45deg and the VA to lie below this line, indicating a shift towards the midpoint of the TBW.

For the training with bright stimuli (Figure 7.1), the TBW for the bright stimuli lie below this line, whereas no training effect is observed for the dim stimuli, as confirmed by the ANOVA which shows a significant interaction between pre vs post-training and stimulus intensity $F(1, 10) = 7.38$, $p = .022$, $\eta p^2 = .43$, but no significant main effect of training $F(1, 10) = 1.21$, $p = .298$, $\eta p^2 = .11$ or of stimulus intensity $F(1, 10) = 3.16$, $p = .106$, $\eta p^2 = .24$. Post-hoc tests revealed that the training with the bright stimuli led to a significant reduction in TBW size for bright stimuli $Z = 1.96$, $p = .025$ (one-tailed) but not for dim stimuli $Z = 0.09$, $p = .465$ (one-tailed). We then tested whether the reduction in TBW size was driven by a criterion shift on the AV, the VA or on both sides (Figure 7.1, row 3). We found a significant shift towards physical simultaneity (SOA = 0) only for the AV criterion $Z = 1.69$, $p = .046$ (one-tailed) but not the VA criterion $Z = 1.07$, $p = .143$ (one-tailed). For the group trained with dim stimuli (Figure 7.2), we find no significant effect of training or stimulus intensity on the TBW size (training: $F(1, 9) = 0.37$, $p = .559$, $\eta p^2 = .04$, stimulus intensity: $F(1, 9) = 0.56$, $p = .474$, $\eta p^2 = .06$) and no significant interaction $F(1, 9) = 0.03$, $p = .861$, $\eta p^2 = .00$.

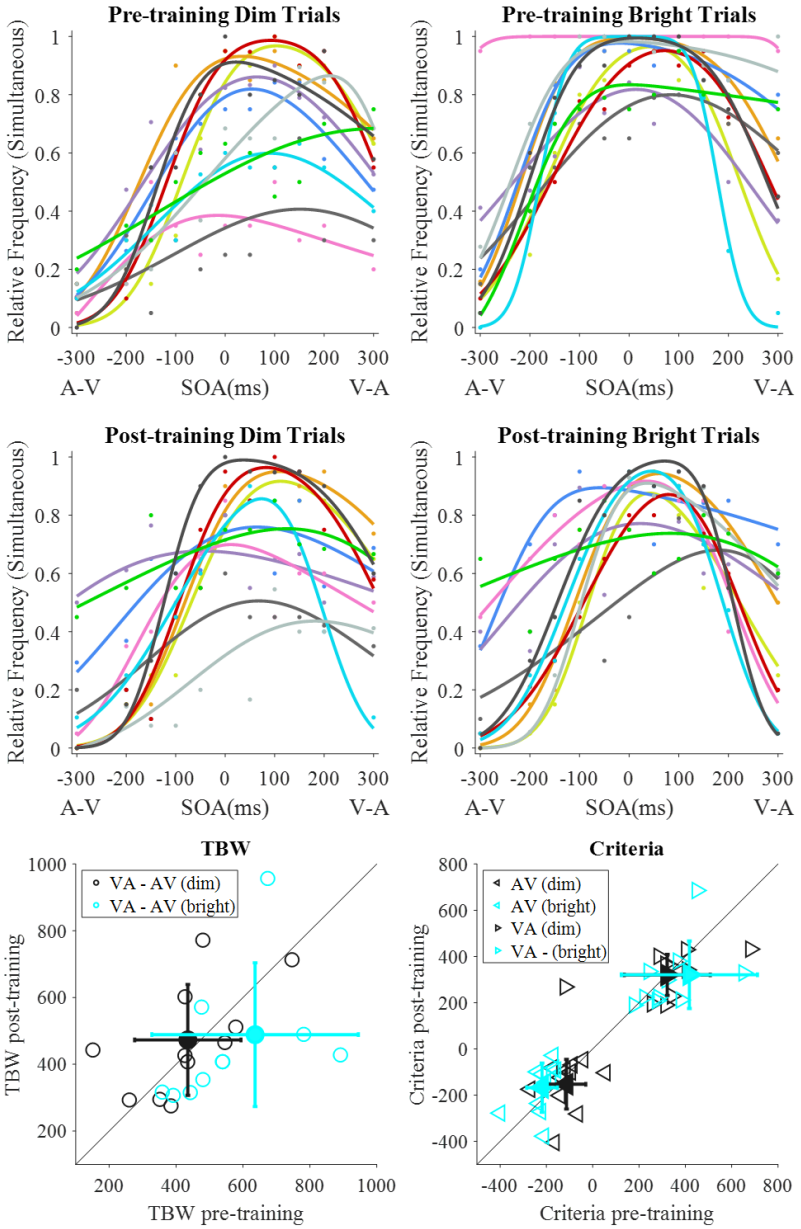


**Figure 7.1.** Bright training group data (n = 11). Rows 1 and 2 contain fits for the pre (row 1) and post (row 2) training data for the dim (left column) and bright (right column) stimuli. Each colour represents an individual observer. Row 3 consists of scatterplots of individual data, alongside means (filled in markers) and standard deviations (error bars). Row 3 (left) represents pre and post-training TBW sizes, calculated as the difference between the two criteria (VA – AV), where data points below the reference line indicate a smaller TBW after training. Row 3 (right) represents the placement of the AV and VA criteria, whereby values above the reference line for the AV criterion and below the line for the VA criterion indicate a shift towards physical simultaneity following training. One data point is not shown in both scatterplots (row 3).

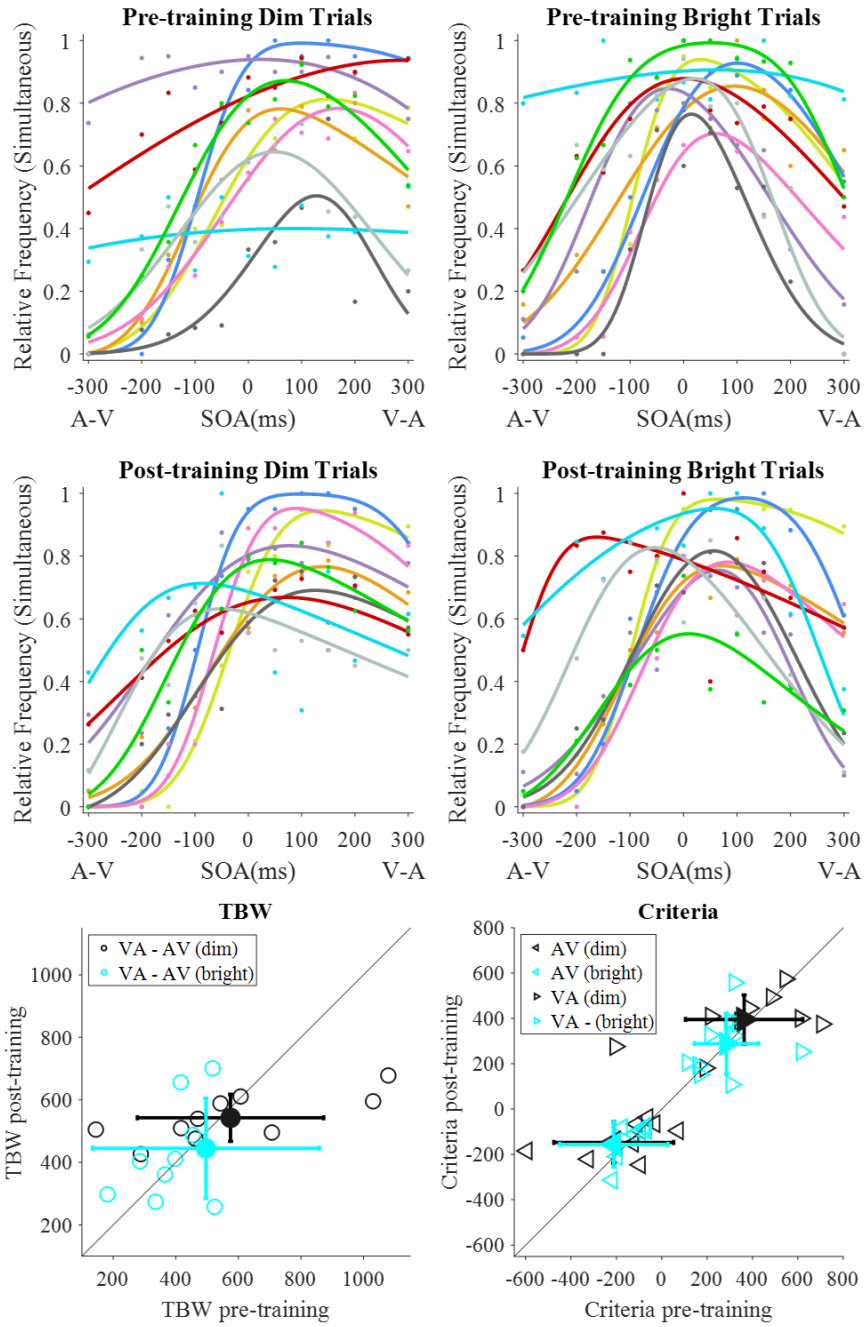


**Figure 7.2.** Dim training group (n = 10). Rows 1 and 2 consist of the individual observer fits. Row 3 contains scatterplots of the pre and post-training TBW size (left) (one data point is not shown) and criteria placement (right).

Our main finding is that training reduces the TBW, but only if the observers were trained with bright stimuli (Figure 7.1). When observers were trained with dim stimuli, no effect of training was observed, neither for dim nor for bright test stimuli (Figure 7.2). This lack of learning could be due to the poor discrimination sensitivity of the SOAs for the dim stimuli as shown in Figure 7.3: the SD for dim stimuli is about twice as large than for bright stimuli. A 2 x 2 x 2 mixed ANOVA, investigating the effect of visual stimulus intensity (dim or bright), leading modality (audio-leading or visual leading) and training group (training with dim or with bright stimuli) on the pre-training SD estimates showed that the main effect of stimulus intensity was approaching significance, with higher SDs for dim stimuli across the two training groups (Table 7.1). Furthermore, there was a significant main effect of leading modality, with higher SDs for visual-leading stimuli. There was no main effect of training group, and importantly, no interaction between training group and leading modality or intensity.

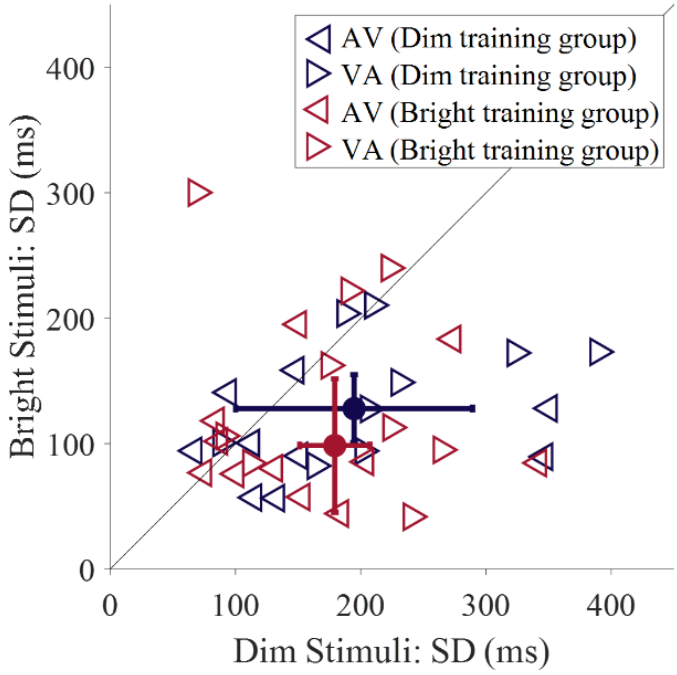


**Figure 7.3.** The SDs of the fitted Gaussians for audio-leading (AV) and visual-leading (VA) pre-training data, across all observers (n = 21). Filled-in circles and error bars represent the median and standard errors, taken across both the AV and VA fits, for the dim (purple) and bright (maroon) training groups. Three data points (> 450) are not shown.

| | Predictor | Sum of Squares | *Df* | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Between-subjects effects | Training group (dim or bright) | 44396 | 1, 19 | 0.35 | .562 | .02 |
| Within-subjects effects | Stimulus Intensity | 165232 | 1, 19 | 4.10 | .057* | .18 |
| | Leading modality (A or V) | 238600 | 1, 19 | 6.85 | .017** | .27 |
| | Intensity * training group | 72354 | 1, 19 | 1.80 | .196 | .09 |
| | Leading modality * training group | 0.51 | 1, 19 | < .001 | .997 | < .001 |
| | Intensity * leading modality | 25094 | 1, 19 | .38 | .545 | .02 |
| | Intensity * leading modality * training group | 92216 | 1, 19 | 1.40 | .252 | .07 |

**Table 7.1.** Summary table for the mixed ANOVA of pre-training SDs. * and ** signify p-values significant at the .1 and .05 levels, respectively.

Finally, we investigated the effect of training, intensity and leading modality on the SDs of the fitted Gaussians. For the bright training group there was a significant main effect of leading modality $F(1, 10) = 12.39$, $p = .006$, $\eta p2 = .55$. Participants had higher SDs to the visual leading stimuli than audio-leading stimuli. The main effects of training and intensity, plus all interactions were non-significant ($p > .1$). As for the dim training group, there was also a main effect of leading modality $F(1, 9) = 20.61$, $p = .001$, $\eta p2 = .70$, with higher SDs for visual leading stimuli. Furthermore, there was a significant main effect of intensity $F(1, 9) = 5.69$, $p = .041$, $\eta p2 = .39$. Observers had higher SDs for dim stimuli than for bright stimuli. All other effects were non-significant ($p > .1$).

### 7.6. Discussion

Previous research has shown that the size of the TBW in audio-visual simultaneity judgements can be reduced through training. Here we tested whether training transferred across different visual signal intensities. While there was no effect of training on perceptual sensitivity (inverse of the SDs of the psychometric functions), individuals trained with bright visual stimuli showed reduced TBW size for bright stimuli after training, an effect that was driven by a criterion shift towards physical simultaneity for audio-leading bimodal stimuli, in line with De Niear, Koo and Wallace, (2016). Importantly, this improvement did not transfer to dim stimuli. The individuals that were trained with dim stimuli, showed no change in TBW, neither for dim nor for bright test stimuli. Our main experimental hypothesis that training is intensity-specific was therefore only partly supported.

The reduction in the TBW when training was performed with bright stimuli is consistent with previous reports (De Niear et al., 2016; Powers et al., 2012, 2009; Zerr et al., 2019). Our analysis also shows that this training effect is consistent with a shift in AV criterion placement, supporting evidence that trial-by-trial feedback can induce criterion shifts (Aberg & Herzog, 2012) which can lead to performance improvements in perceptual tasks (Herzog et al., 2006). The stimulus specificity of this criterion shift is consistent with perceptual learning as a mechanism underlying the performance improvement (Herzog et al., 2006).

Individuals trained with the dim visual stimuli showed *no* significant change in TBW size with training. Our data show that participants in both training groups had lower initial (pre-training) sensitivity (for dim stimuli than for bright stimuli. The reliability of stimulus-feedback combinations is an important aspect in training induced reductions in audio-visual TBW size (De Niear, Noel, & Wallace, 2017), and in subjective confidence in perceptual judgements (Boldt, de Gardelle, & Yeung, 2017). We speculate that a relative unreliability

in the perception of audio-visual offsets for low-intensity stimuli, may have impeded the effectiveness of the dim intensity training; if participants are less likely to consistently perceive simultaneity, or non-simultaneity, at a given SOA, then more perceptual variability is introduced when training these individuals to provide a specific response at said offset. Additionally, consistent with previous research, observers had much lower SDs for audio-leading in contrast to visual leading stimuli (Cecere et al., 2016; van Eijk, Kohlrausch, Juola, & van de Par, 2008). It has been argued that this is due to the low-level attentional effects of auditory signals, which alert the visual system to an upcoming visual stimulus (Thorne & Debener, 2014), resulting in higher sensitivities for audio-leading offsets (Cecere et al., 2016).

The ability to accurately integrate signals from different sensory modalities has potential wide-ranging effects on daily living. Increases in TBW size have been associated with increased fall risk in older age (Setti et al., 2011), and poorer performance on speech perception tasks (Conrey & Pisoni, 2006; Stevenson et al., 2018). Training healthy individuals at an audio-visual SJ task led to lasting improvements in speech perception (Zerr et al., 2019). Training with the specific aim to reduce the TBW has consequently been proposed as an intervention for those with multisensory deficiencies (Foss-Feig et al., 2010; Powers et al., 2009; Setti et al., 2014; Stevenson et al., 2017; Wallace & Stevenson, 2014). Our results highlight that, for judgments of perceived simultaneity, perceptual learning is specific for high visual intensity stimuli. Such a finding potentially highlights limitations in proposed interventions; if training is specific to a visual intensity, or intensities, then training-induced improvements may not be applicable in dynamic, real-world settings. The specificity of perceptual learning increases with training (Jeter et al., 2010), and, to be useful across a range of conditions, a careful balance between specificity and generalisation has to be struck. This consideration may also explain conflicting results, such as the

generalisation of unimodal TOJ training to bimodal stimuli (e.g., Stevenson et al., (2013) vs. Zerr et al., (2019)).

Previous research shows that perceptual learning is possible at near threshold stimuli (e.g. Andersen, Ni, Bower, & Watanabe, 2010; Cong, Wang, Yu, & Zhang, 2016). While our results could potentially highlight a task specific limitation, where the SJ task is untrainable at low visual intensities, we argue that a more parsimonious explanation, is that in our paradigm, training is non-transferable from a high visual intensity to a low visual intensity. However, we should not discount the former, and so further research should be conducted with various visual intensities.

In conclusion, the temporal window within which audio-visual stimuli are perceived as simultaneous can be narrowed by training, but the effect of training does not transfer from the high to low visual intensity stimuli. This improvement is driven by a criterion shift for auditory-leading bimodal stimuli towards physical simultaneity.

### 7.7. Chapter summary

The current chapter investigated the generalisability of training induced reductions in TBW size across stimuli with different visual intensities. The results showed that the plasticity in the TBW was a result of a shift in the auditory-leading criterion, but that there was no increase in perceptual sensitivity following training. Furthermore, the improvements following training with bright stimuli, were not generalised to the post-training dim trials. The stimulus specificity of this training highlights potential limitations of SJ training, which will be discussed in more detail in the following chapter. The following chapter will also discuss the results of this thesis in relation to previous literature and will discuss potential directions for future research in light of the findings presented in this thesis.

# 8.General Discussion

## 8.1. Chapter Introduction

This section briefly discusses the literature and methodology discussed in chapters 1-3, and then summarises the results and conclusions of each experimental chapter (4-7). Following the chapter summaries, the results of the current thesis are discussed in relation to previous research, with a specific focus on the effect of visual intensity on perceived relative timing. It discusses the plasticity of perceived simultaneity, followed by the role of processing latencies and then decisional processes in perceived relative simultaneity. It addresses how the current data addresses various unresolved gaps in the literature and discusses the future directions that should be taken in this field of study.

## 8.2. Summary of individual chapters

The current thesis focused on the intensity dependency and plasticity of perceptual judgments on relative timing, with a specific focus on the behavioural and neural correlates of the simultaneity and temporal order judgement tasks. **Chapter 1** provided a basic summary of the processes underlying the perception of audio and visual stimuli, and highlighted the separate processing pathways of the two modalities. It then discussed the neural processes that are specific to the perception of bimodal audio-visual stimuli, and the main principles which dictate whether these stimuli are integrated (see Meredith & Stein, 1986). Finally, this chapter described the effect of stimulus intensity on neural firing rates, and the resultant effect on sensory processing latencies.

**Chapter 2** discussed the SJ and TOJ tasks, and what their estimates can tell us about audio-visual processing and how they have been used to study various clinical disorders. Various models that have attempted to explain the processes underlying the tasks were discussed,

with a particular focus on the MCD model. This chapter also described the TBW, the plasticity of this window and the various disorders where a larger than typical TBW is often observed. Finally, the prolonged and rapid recalibration of perceived simultaneity was discussed.

**Chapter 3** summarised the tailormade equipment, the use of EEG, and the various psychophysical paradigms utilised throughout the presented thesis. It summarised Yarrow et al.'s (2011) fitting method and the various stages of the MCD model, both using hypothetical data.

**Chapter 4** was the first experimental chapter. Here, participants completed the SJ, TOJ and simple RT tasks at two visual intensities which were varied across blocks. The results were consistent with previous findings (e.g., Van Eijk et al., 2008), with no significant correlation found between the PSSs estimated using the SJ and TOJ tasks, for both the high and low intensity trials. Further, the previously reported effect of visual intensity on the PSS was replicated (e.g., Leone & McCourt, 2015), with a PSS shift towards the visual leading SOAs for the dimmer visual intensity. The effect of intensity on the PSS contrasted with the corresponding effect on facilitative multisensory integration, where race model violations were observed only at physical simultaneity, regardless of stimulus intensity. This finding was consistent with Leone and McCourt (2015), who concluded that differences in central arrival latencies caused by stimulus intensity, are compensated for in the 'action' (i.e. the RT) but not 'perception' (i.e. SJ and TOJ) tasks. While the experiments in this thesis cannot confirm the hypothesised tasks-specific process, it at least supports the previous behavioural data (n = 5) with a much larger sample (n = 38).

The experiments in **chapter 5** built upon the previous chapter, as several visual intensities were presented to observers, in the SJ, TOJ and simple RT tasks. The simple RT data were used to estimate the parameters of an intensity-dependent pre-processing stage of the

MCD model. The results of the model comparison suggested that the effect of visual intensity on the PSS is accounted for entirely by intensity-dependent processing latencies, and furthermore, that the decisional stage is *independent* of stimulus intensity.

Previous research has shown that the PSS rapidly recalibrates based on the SOA of a preceding trial (Noel et al., 2016; Van der Burg et al., 2013, 2018; Van der Burg & Goodbourn, 2015). Since chapter 5 proposed that the effect of visual intensity on the PSS is due to intensity-dependent processing latencies, **chapter 6** investigated whether the PSS would rapidly recalibrate based on the intensity (and therefore relative central arrival times) of the preceding visual stimulus. The results highlighted an effect of both the SOA (SJ and TOJ tasks) and intensity (TOJ task only) of the preceding trial. Corresponding ERP data revealed an effect of preceding intensity in both the SJ and TOJ tasks, though this was evident at only one of the three SOAs used for each task, and with relatively small effect sizes when tested statistically. It was proposed (albeit tentatively) that rapid recalibration to preceding intensity is driven by a process that is consistent with the previously reported rapid recalibration to preceding SOA. Since Chapters 4 and 5 reported uncorrelated PSSs, which have been argued to reflect differences in integration mechanisms (Zampini et al., 2003), or differences in decisional processes (e.g., Parise & Ernst, 2016), Chapter 6 also aimed to assess how the temporal dynamics of the neural processes underlying the tasks differ. The analysis revealed only late differences, indicative of variation at the decisional stage.

Following the plasticity in perceived simultaneity, highlighted in chapter 6, **chapter 7** focused on the plasticity of the TBW, by assessing whether training induced reductions in TBW size are generalisable across stimuli of differing visual intensities. The analysis revealed that training induced improvements were not generalisable from the bright to the dim trials, emphasising the stimulus specificity of this effect. By using the fitting procedure

from Yarrow et al. (2011), the effect of training on perceptual sensitivity and the placement of decisional criteria was disassociated. The observed training induced improvements for bright trials were attributed to a criterion shift, rather than an increase in perceptual sensitivity.

### 8.3. The plasticity of perceived simultaneity

The plasticity of the perception of simultaneity was demonstrated in this thesis, both in terms of a rapid plasticity, in response to the temporal offset and visual intensity of an immediately preceding trial, and in terms of a plasticity in response to a prolonged training procedure. This plasticity is discussed in this section.

An audio-visual event at a certain distance to the observer must be made up of very specific relative intensities for the respective signals to reach multisensory areas of the brain at simultaneous latencies. Of course, for real-world stimuli, these perfect conditions are often not met, and so some of this variability in central arrival times is accounted for by the TBW. The TBW has been shown to be highly plastic in response to training procedures (e.g., Powers et al., 2009), and training individuals to reduce the size of their TBW has been proposed as an intervention for those with multisensory deficiencies (Powers et al., 2009; Setti et al., 2014; Wallace & Stevenson, 2014).

Chapter 7 reaffirmed the TBWs plasticity in response to training, but also highlighted the stimulus specificity of this effect. In doing so, it exposed two potentially important problems with the previously proposed interventions. Since relative intensities constantly fluctuate in the real world, it is essential that any improvements can be generalised across a wide range of stimuli; the findings presented in chapter 7 potentially raises questions on whether this is the case for the changes seen after SJ training. Moreover, improvements in the perceptual system's ability to discriminate the physical offsets between audio-visual stimuli (which would be evident through increased perceptual sensitivity estimates) may

translate to a reduction in the incorrect binding of audio-visual signals (e.g., Stevenson et al., 2018), thereby reducing perceptual errors. The fact that this was not observed in the presented data therefore highlights the second potential limitation of the proposed interventions, at least in terms of the training paradigm used in chapter 7.

Despite there being no increase in perceptual sensitivity, the SJ training led to a significant criterion shift. It should be noted however, that comparable criteria shifts can also be induced through simple variations in instructions, such as by simply asking observers to be "more conservative" with their responses in the SJ task (Yarrow & Roseboom, 2017). It is therefore clear that if the aim of an experiment is to manipulate the decisional criteria underlying SJs, the lengthy training procedure used in chapter 7 is far from the most efficient method to achieve this.

It must be emphasised that the data presented in chapter 7 should be interpreted with some caution, however. For several participants (most notably in the bright training group), the estimates of decisional criteria were extrapolated from the curve fitted to the data, and lie outside the tested range of SOAs. It is therefore possible that the criteria estimates are unreliable, since a small change in the relative frequency of 'simultaneous' responses could lead to a large shift in criteria estimates when the data is fitted, especially in those with abnormally large TBWs. Since the TBWs reported for some individuals in this chapter are larger than what is reported in other experiments which incorporate a wider range of SOAs (e.g. De Niear et al., 2016), it is possible that the relative frequency of 'simultaneous' responses would reduce to zero at SOAs much smaller than the extrapolated estimates reported in this chapter.

Because past research into the effect of SJ training on TBW size has confounded perceptual sensitivity with criterion placement (e.g., Powers et al., 2009), it would be beneficial for these data sets to be reanalysed, in order to assess whether the training procedures led to

a true increase in perceptual sensitivity, and therefore further assess the suitability of SJ training as an intervention for individuals with multisensory deficiencies.

### 8.3.1. The role of TBW and PSS plasticity, in accounting and compensating for intensity-dependent processing latencies

Based on the results of chapter 7, the author cautiously proposes that the decisional criteria required to make SJs are placed and adjusted, at least partially independently, for stimuli with different intensities. This finding is consistent with the idea that it is possible to simultaneously hold differentially placed decisional criteria for different stimuli in perceptual tasks (Banerjee, Grover, Ganesh, & Sridharan, 2017). Such a process would allow for flexibility in judgements on the relative timing of audio-visual stimuli, whose characteristics impact the processing delay of their comprising modalities. For example, lower intensity visual stimuli will arrive at multisensory regions of the brain later than high intensity visual stimuli, and the aforementioned flexibility would allow for these intensity dependent differences in central arrival times to be taken into account when judging audio-visual simultaneity.

While holding differentially placed decisional criteria for stimuli of different visual intensities is theorised here to be important in accounting for intensity-dependent differences in relative processing latencies, the results of chapter 6 potentially reflect a mechanism to *compensate* for these intensity-dependent differences. Here, a plasticity of the PSS was observed in the TOJ task, which corrected for the visual-intensity dependent processing latencies of the preceding trial. The author therefore posits that rapid recalibration exists to compensate (at least in part), for the rapidly changing stimulus intensities of an individual's surrounding environment.

## 8.4. The role of processing latencies in the perception of simultaneity

Reaction times are used to provide an insight into the processing latencies of the visual system (Pins & Bonnet, 1997), whereas the PSS is used as an estimate of *relative* processing

latencies (e.g., Barnett-Cowan & Harris, 2009). The simple RT data reported in Chapters 5 and 6, provide important insights into the role of intensity-dependent processing latencies in the perception of simultaneity, and in rapid recalibration. This role of processing latencies is discussed in this section.

The experiments presented in chapters 4, 5 and 6 replicated the previously reported effect of visual intensity on the PSS. While it might be expected that visual intensity manipulations would have consistent effect on simple RTs and on the PSS (assuming the auditory intensity used when estimating the PSS is held constant), previous research has shown that the effect of visual intensity on simple unimodal RTs, can be up to double the corresponding effect on the PSS (Jaśkowski, 1992; Menendez, & Lit, 1983). It has been suggested that this disparity may be driven by an effect of attentional bias on judgements of relative timing which is non-existent for simple RTs (Jaśkowski, 1992), although another alternative could be related to the average perceived offset in a block of trials (Miyazaki et al., 2006), which can vary due to relative stimulus intensity (see section 2.9). Importantly, the potential of either possibility impacting responses, exists exclusively when stimulus intensity is manipulated across blocks (e.g., Jaśkowski, 1992; Menendez, & Lit, 1983).

In contrast to previous research, the experiments presented in Chapter 5 involved the presentation of interleaved stimulus intensities within blocks, and a highly similar effect of intensity on simple RTs and the PSS, was evident in the data. The results are therefore consistent with the idea that the intensity dependency of simple RTs, and of the PSSs of both the SJ and TOJ tasks, are driven by shared intensity-dependent processing latencies.

#### 8.4.1. Rapid recalibration may contribute to the differences between PSS and RT estimates observed previously.

The effect of rapid recalibration to preceding visual intensity (reported in chapter 6), could contribute to the previously reported discrepancy in the intensity-dependent PSS shift in the TOJ task, and the corresponding effect of intensity on unimodal RTs. Based on the

results of chapter 6, if visual intensity is held constant within a block, then the PSS will rapidly recalibrate towards the visual-leading SOAs following higher intensity stimuli, and towards the auditory-leading SOAs following lower intensity stimuli. If visual intensity is manipulated only across blocks, then the impact of this rapid recalibration would partially reduce the effect of intensity on PSS, relative to the corresponding effect of intensity on RTs. For example, the PSS is typically estimated to be shifted towards the auditory-leading SOAs for brighter stimuli, relative to dim. However, if this stimulus is preceded by another bright stimulus, then the perception of simultaneity would have shifted towards the *visual*-leading SOAs, thereby lessening the intensity dependency of the PSS.

Similarly to previous research, Harris et al. (2010) reported that PSS shifts due to reductions in visual processing latency (achieved by placing light attenuating eyewear on the observer in a TOJ task with a block design) were not consistent with the corresponding effect on unimodal RTs. Hence, they concluded that a mechanism exists to compensate for the variability in visual processing latencies. Rapid recalibration, based on the intensity of the preceding trial is a compensatory mechanism that could, in theory, explain this inconsistency in Harris et al.'s (2010) findings.

### 8.4.2. Rapid recalibration is not driven by changes in early processing latencies

While the process of rapid recalibration could potentially contribute to the inconsistent effect of intensity on simple RTs and the PSS, rapid recalibration itself cannot be explained by changes in early processing latencies. In chapter 6, the analyses revealed no effect of the preceding trial's intensity on simple RTs or ERP latencies. While the inconsistency between the effect of the preceding trial on the PSS and on RTs, is not entirely unexpected, given that shifts in the PSS are not necessarily accompanied by changes in early processing latencies (McDonald et al., 2005), the results highlight an interesting contrast to the

recalibration observed following prolonged exposure to temporally offset bimodal stimuli, which *is* accompanied by changes in simple RTs (Di Luca et al., 2009; Navarra et al., 2009).

When taken together, the results of Chapters 5 and 6 suggest that the PSS (i.e. of the current trial) is dependent on visual processing latencies, yet visual intensity manipulations have no impact on the processing latency of the *subsequent* trial. The latter suggests that *rapid* recalibration, both due to physical temporal offsets and intensity-dependent processing latencies, is compensated for via a different process to the process which proposedly underlies prolonged recalibration (Di Luca et al., 2009; Navarra et al., 2009). Rapid and prolonged recalibration have previously been shown to operate on different time scales (Van Der Burg et al., 2015), and the data in chapter 6 highlights further differences between these processes.

Why would rapid changes in central arrival times be compensated for via separate process to prolonged offsets? A tentative explanation is that processing latencies may adapt to prolonged offset stimuli, such as an audio-visual signal from a source a large distance from the observer, to avoid the persistent demand for the sensory-decisional process underlying rapid recalibration. Such a process requires repetitive offset trials, and therefore a mechanistically separate 'rapid' recalibration process may exist to swiftly compensate for differing central arrival times (such as those caused by intensity) before the proposed adjustment of processing latencies takes place.

### 8.5. The role of decisional processes in the perception of simultaneity

The uncorrelated PSS estimates of the SJ and TOJ tasks is accounted for in the MCD model by the $\overline{MCD_{CORR}}$ and $\overline{MCD_{LAG}}$ computations, which are weighted differently for the two tasks. Furthermore, the lack of an effect of the preceding trial's properties on the simple RTs of the current trial suggests that rapid recalibration is also driven by a decisional

process. This section discusses the role of decisional processes in the perception of simultaneity, and highlights the contribution of the presented ERP findings to this debate.

### 8.5.1. Variability in the SJ and TOJ tasks at the decisional stage

The uncorrelated PSSs previously reported from the SJ and TOJ tasks (Basharat et al., 2018; Binder, 2015; Linares & Holcombe, 2014; Love et al., 2013; Van Eijk et al., 2008; Vatakis et al., 2008) has raised the question of whether the two tasks are driven by separate neural mechanisms (e.g., Zampini et al., 2003; Love et al., 2013). Since the two tasks are often used in combination to study various disorders (e.g., Capa et al., 2014; Regener et al., 2014; Scarpina et al., 2016), it is vital that an answer to this question is sought. Consistent with past findings, the behavioural data reported in chapters 4, 5 and 6 consistently replicated the lack of correlation between the PSSs of the two tasks. While the behavioural experiments in this thesis highlight the robustness of this finding, this alone cannot confirm that the tasks are driven by separate neural mechanisms.

EEG has a very high temporal resolution (e.g., Burle et al., 2015), and through the analysis of ERPs, it is possible to distinguish whether perceptual tasks differ at an early sensory processing stage (typically evident as early differences in the ERP) (e.g., Ahmadi, McDevitt, Silver, & Mednick, 2018; Gohil, Hahne, & Beste, 2016) or at later decisional stages (Ahmadi et al., 2018; Gohil et al., 2016). For example, if the SJ task draws upon multisensory binding mechanisms (which are typically shown to effect early latency ERPs, see section 3.13.1), as opposed to a separate mechanism involved in the perception of temporal order (Zampini et al., 2003), then early differences in the ERPs of the two tasks would likely be expected.

ERPs were therefore used to pinpoint the timing of potential differences in the neural processes underlying the SJ and TOJ tasks, which could not be done using behavioural methods alone. The results of the ERP analysis presented in chapter 6 revealed that the tasks differed only late in the ERP (> 400ms). These results are highly consistent with the

MCD model (chapter 5), both in terms of the identical early processing for the auditory and visual signals required to make SJs and TOJs, alongside the late differences in the ERPs, perhaps indicative of the variable utilisation of these signals at the decisional stage. This variability at the late processing stage may have relevance for researchers who use the SJ and TOJ tasks in clinical settings, by highlighting the need to understand the exact influence of these varying decisional processes on the conflicting estimates produced from the SJ and TOJ tasks.

### 8.5.2. The role of decisional processes in rapid recalibration

If the SJ and TOJ tasks share an intensity-dependent early processing stage, why would the intensity of a preceding visual stimulus have differing impacts, dependent on task, on the PSS of the following trial? The answer to this may be apparent when considering that the effect of preceding visual intensity was only evident late (>300ms) in the ERP, across both the SJ and TOJ tasks, consistent with the possibility that this effect is driven by later, decisional processes (Kelly & O'Connell, 2013; O'Connell et al., 2012; Simon et al., 2017; Tagliabue et al., 2019). However, further research would be required to assess the robustness of this effect (see section 6.7).

Linking back to the MCD model, since preceding visual intensity had no impact on simple RTs or ERP latencies, it would be expected that the $\overline{MCD}_{CORR}$ and $\overline{MCD}_{LAG}$ outputs would be identical across trials preceded by bright and dim stimuli (and across both the SJ and TOJ tasks). Consequently, the observed PSS shift would be driven by a later, task-dependent decisional effect which modulates how the two model outputs are utilised at the decisional stage.

## 8.6. Contrasting early and late effects

The above conclusions highlight an interesting juxtaposition. Taken together, the ERP findings from chapter 6, and the largely consistent effect of visual intensity on simple RTs

and the PSS in chapter 5, suggest that the intensity-dependency of perceived simultaneity is driven by visual processing latencies. During this early sensory processing stage, auditory and visual signals are processed identically, regardless of whether the SJ or TOJ task is being performed. This is in stark contrast to the task-dependent PSS differences observed across tasks, which are driven entirely by differences in the way these shared sensory representations are utilised at the decisional stage (Parise & Ernst, 2016). Likewise, the plasticity in perceived simultaneity, both in terms of rapid recalibration, and in terms of the training induced changes in TBW size, can also be explained by decisional processes which occur after the (intensity-dependent) relative arrivals of the audio-visual signals at multisensory areas of the brain.

### 8.7. Methodology critique

A potential weakness of the experimental chapters presented in this thesis is the consistent use of simple stimuli with low ecological validity across the various tasks. The flash-beep trials lack the complexity of stimuli sometimes used for the SJ and TOJ tasks. For example, they lack the predictive information given by stimuli with continuous motion, such as an audio-visual video of a bouncing ball (van Eijk et al., 2008) or a hammer hitting an object (Stevenson et al., 2014), and they also lack cues on distance, which are given by varying the distance of the equipment (which produces the stimuli) to the observer, in a room that is illuminated enough for this equipment to be visible (Harris et al., 2010). Research has highlighted differences between various estimates from simple and more complex stimuli. For example, the TBW is estimated to be larger for more complex stimuli (Stevenson et al., 2014; Stevenson & Wallace, 2013), but the effect of preceding SOA on the PSS can be smaller (Noel et al., 2017), which therefore raises questions on the applicability of the findings reported in the current thesis to the perception of real-world stimuli.

An additional drawback of using the simple flash-beep stimuli has specific relevance to chapter 6. Here, the effect of preceding visual intensity on VEP latency was assessed, however, it was not possible to assess the intensity-dependency of the VEP latency (i.e. of the current trial), since the dim visual stimulus did not produce a reliable VEP across observers. Furthermore, the analysis for the effect of preceding intensity on the VEP latency of the 'test' intensity trials was restricted to a time window that was later than what is typically used to assess VEP latencies (e.g., Sharma, Joshi, Singh, & Kumar, 2015), since the stimuli did not evoke the typical visual P1 and N1 components (at least not at the latencies where they are typically expected to occur). This is likely due to the small, relatively low intensity stimuli presented to observers, as both VEP amplitude (Brannan, Solan, Ficarra, & Ong, 1998) and latency (Carrillo-de-la-Pena et al., 1999) are dependent on intensity. Other stimuli could have been used to elicit a more reliable VEP, such as pattern-reversal checkerboard stimuli (Kovarski et al., 2019), potentially allowing for more informative analyses to be run on this data set.

There were several benefits of using simple stimuli however, that were important for assessing the hypotheses in this thesis. For example, it allowed the researcher a high level of control and precision when presenting audio-visual trials, without being limited to the refresh rate of a computer monitor (which would be required for the presentation of other, more complex stimuli). Furthermore, the use of LEDs meant that the background luminance inside the soundproof booth could be minimised, which was important when investigating the effect of intensity on responses; again, this would not be the case when presenting a stimulus in the centre of a computer monitor.

A final drawback of the experimental designs used in the current thesis stems from the SOAs of the audio-visual trials presented to observers (either -200 AV to 200 VA, or -300 AV to 300 VA). The SJ and TOJ tasks can be subjectively difficult, and it was relatively common

for participants to be rejected from experiments due to poor performance in the practice blocks of these tasks. The selected range of SOAs aimed to provide a reasonable level of difficulty for observers, while not being numerous enough that experiments became uncomfortably long. An alternative would have been to run more blocks over more testing sessions, therefore allowing more trials to be presented over a larger range of SOAs (while still ensuring small intervals between these SOAs). This would have ensured that the largest presented audio-visual offsets were more easily perceptible to observers, therefore reducing the number who were rejected due to poor practice block performance (e.g., section 5.5.3), and allowing the samples to be more representative of the general population, rather than just the best performing observers.

### 8.8. Directions for the future

A central focus of this thesis was on the difference between the SJ and TOJ tasks. The results presented in chapter 6 showed only late differences in the ERPs of the two tasks, consistent with the theory that the tasks differ only at the decisional stage. Previous fMRI experiments have shown both unique (Binder, 2015) or entirely overlapping (Love et al., 2018) neural structures active during the tasks, but a common finding across these experiments is the higher BOLD signal elicited by the TOJ, over the SJ task. Since BOLD signal is highly dependent on task difficulty (Culham et al., 2001; Gould et al., 2003), it is unclear as to which (if any) of the reported differences were driven by the subjectively higher difficulty of the TOJ task (Love et al., 2013). Future research should focus on balancing the difficulty of the two tasks, either through training, or the use of different SOAs which show an equal correct response rate across tasks. This would aim to minimise the influence of differential task difficulty on neural activation, therefore isolating any neural structures that could be specific to the processing underlying the SJ or TOJ task.

Finally, the size of the TBW varies based on stimulus complexity, with increasing TBWs typically estimated for more complex stimuli (Stevenson et al., 2014; Stevenson & Wallace, 2013). An important focus of future research should be on whether SJ training induced reductions in TBW size using audio-visual speech stimuli, can generalise across different speakers. It should be noted that performance improvements in SJ training correlate with pre-training TBW size (Powers et al., 2009), and therefore it may be expected that training with more complex stimuli (which have been shown to be integrated over a larger TBW) will lead to greater improvements. Since previous perceptual learning research has shown that greater performance improvements are associated with higher levels of learning generalisation across stimuli of differing visual intensities (Lengyel & Fiser, 2019), it could be possible that training observer with audio-visual speech stimuli will lead to performance improvements that can be generalised across different speakers. Such a finding would provide strong evidence in support of the efficacy of previously proposed training-based interventions, with regards to real-world stimuli that are associated with perceptual errors in those with multisensory deficiencies (Stevenson et al., 2018).

# 9. References


Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3), 1–11. https://doi.org/10.1167/12.3.3

Adams, W. J. (2016). The Development of Audio-Visual Integration for Temporal Judgements. *PLOS Computational Biology*, *12*(4), e1004865. https://doi.org/10.1371/journal.pcbi.1004865

Adler, G., & Adler, J. (1989). Influence of stimulus intensity on AEP components in the 80- to 200-millisecond latency range. *International Journal of Audiology*, *28*(6), 316–324. https://doi.org/10.3109/00206098909081638

Ahmadi, M., McDevitt, E. A., Silver, M. A., & Mednick, S. C. (2018). Perceptual learning induces changes in early and late visual evoked potentials. *Vision Research*, *152*, 101–109. https://doi.org/10.1016/j.visres.2017.08.008

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

Allen, W. F. (1927). Location in the spinal cord of pathways which conduct impulses from the cerebrum and superior colliculus, affecting respiration. *The Journal of Comparative Neurology*, *43*(3), 451–515. https://doi.org/10.1002/cne.900430305

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267–278.

https://doi.org/10.1016/S1364-6613(00)01501-1

Allman, B. L., Bittencourt-Navarrete, R. E., Keniston, L. P., Medina, A. E., Wang, M. Y., & Meredith, M. A. (2008). Do Cross-Modal Projections Always Result in Multisensory Integration? *Cerebral Cortex*, *18*(9), 2066–2076. https://doi.org/10.1093/cercor/bhm230

Alpert, G. F., Hein, G., Tsai, N., Naumer, M. J., & Knight, R. T. (2008). Temporal characteristics of audiovisual information processing. *Journal of Neuroscience*, *28*(20), 5344–5349. https://doi.org/10.1523/JNEUROSCI.5039-07.2008

Andersen, G. J., Ni, R., Bower, J. D., & Watanabe, T. (2010). Perceptual learning, aging, and improved visual performance in early stages of visual processing. *Journal of Vision*, *10*(13), 1–13. https://doi.org/10.1167/10.13.4

Anderson, D. R., & Burnham, K. P. (2002). Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management*, *66*(3), 912-918. https://doi.org/10.2307/3803155

Angus-Leppan, H. (2008). Diagnosing epilepsy in neurology clinics: A prospective study. *Seizure*, *17*(5), 431–436. https://doi.org/10.1016/j.seizure.2007.12.010

Antunes, F. M., Nelken, I., Covey, E., & Malmierca, M. S. (2010). Stimulus-specific adaptation in the auditory thalamus of the anesthetized rat. *PLoS ONE*, *5*(11), e14071–e14071. https://doi.org/10.1371/journal.pone.0014071

Arnold, J. E., & Watson, D. G. (2015). Synthesising meaning and processing approaches to prosody: performance matters. *Language, Cognition and Neuroscience*, *30*(1–2), 88–102. https://doi.org/10.1080/01690965.2013.840733

Banerjee, S., Grover, S., Ganesh, S., & Sridharan, D. (2017). Sensory and decisional components of endogenous attention are dissociable. *Journal of Neurophysiology*,

*122*(4), 1538-1554. https://doi.org/10.1101/231613

Bang, J. W., & Rahnev, D. (2017). Stimulus expectation alters decision criterion but not sensory signal in perceptual decision making. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-16885-2

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

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*(3), 377–391. https://doi.org/10.1162/0898929053279586

Bartlett, N. R., & Macleod, S. (1954). Effect of flash and field luminance upon human reaction time. *Journal of the Optical Society of America*, *44*(4), 306–311. https://doi.org/10.1364/JOSA.44.000306

Barutchu, A., Danaher, J., Crewther, S. G., Innes-Brown, H., Shivdasani, M. N., & Paolini, A. G. (2010). Audiovisual integration in noise by children and adults. *Journal of Experimental Child Psychology*, *105*(1–2), 38–50. https://doi.org/10.1016/j.jecp.2009.08.005

Basharat, A., Adams, M. S., Staines, W. R., & Barnett-Cowan, M. (2018). Simultaneity and Temporal Order Judgments Are Coded Differently and Change With Age: An Event-Related Potential Study. *Frontiers in Integrative Neuroscience*, *12*. https://doi.org/10.3389/fnint.2018.00015

Bean, B. P. (2007). The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, *8*, 451–465. https://doi.org/10.1038/nrn2148

Bedard, G., & Barnett-Cowan, M. (2016). Impaired timing of audiovisual events in the elderly. *Experimental Brain Research*, *234*(1), 331–340. https://doi.org/10.1007/s00221-015-4466-7

Behne, D. M., Sorati, M., & Alm, M. (2017). Perceived Audiovisual Simultaneity in Speech by Musicians and Non-musicians: Preliminary Behavioral and Event-Related Potential (ERP) Findings. *AVSP*, 100–104. https://doi.org/10.21437/AVSP.2017-19

Bennett, C., Miller, M., & Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage*, *47*, S125. https://doi.org/10.1016/s1053-8119(09)71202-9

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology Biofeedback*, *42*, 247–255. https://doi.org/10.1007/s10484-017-9371-3

Bernasconi, F., Grivel, J., Murray, M. M., & Spierer, L. (2010). Plastic brain mechanisms for attaining auditory temporal order judgment proficiency. *NeuroImage*, *50*(3), 1271–1279. https://doi.org/10.1016/j.neuroimage.2010.01.016

Besle, J., Fort, A., & Giard, M. H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing*, *5*(3), 189–192. https://doi.org/10.1007/s10339-004-0026-y

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

Boenke, L. T., Deliano, M., & Ohl, F. W. (2009). Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. *Experimental Brain Research*, *198*(2–3), 233–244. https://doi.org/10.1007/s00221-009-1917-z

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520–1531. https://doi.org/10.1037/xhp0000404

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

Bonmassar, G., Anami, K., Ives, J., & Belliveau, J. W. (1999). Visual evoked potential (VEP) measured by simultaneous 64-channel EEG and 3T fMRI. *NeuroReport*, *10*(9), 1893–1897. https://doi.org/10.1097/00001756-199906230-00018

Brainard, D. H., Cottaris, N. P., & Radonjić, A. (2018). The perception of colour and material in naturalistic tasks. *Interface Focus*, *8*(4), 20180012. https://doi.org/10.1098/rsfs.2018.0012

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex*, *23*(6), 1329–1341. https://doi.org/10.1093/cercor/bhs109

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early

adolescence: A high-density electrical mapping study. *Cerebral Cortex*, *21*(5), 1042–1055. https://doi.org/10.1093/cercor/bhq170

Brannan, J. R., Solan, H. A., Ficarra, A. P., & Ong, E. (1998). Effect of luminance on visual evoked potential amplitudes in normal and disabled readers. *Optometry and Vision Science*, *75*(4), 279–283. https://doi.org/10.1097/00006324-199804000-00025

Brelén, M. E., Vince, V., Gérard, B., Veraart, C., & Delbeke, J. (2010). Measurement of evoked potentials after electrical stimulation of the human optic nerve. *Investigative Ophthalmology and Visual Science*, *51*(10), 5351–5355. https://doi.org/10.1167/iovs.09-4346

Brenner, E., & Smeets, J. B. J. (2019). How Can You Best Measure Reaction Times? *Journal of Motor Behavior*, *51*(5), 486–495. https://doi.org/10.1080/00222895.2018.1518311

Brisson, B., Robitaille, N., & Jolicýur, P. (2007). Stimulus intensity affects the latency but not the amplitude of the N2pc. *Neuroreport, 18*(15), 7–10. https://doi.org/10.1097/WNR.0b013e3282f0b559.

Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology*, *97*(3), 210–220. https://doi.org/10.1016/j.ijpsycho.2015.05.004

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

Butera, I. M., Stevenson, R. A., Mangus, B. D., Woynaroski, T. G., Gifford, R. H., & Wallace, M. T. (2018). Audiovisual Temporal Processing in Postlingually Deafened Adults with Cochlear Implants. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-

29598-x

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, *10*(12), 2619–2623. https://doi.org/10.1097/00001756-199908200-00033

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, *14*(2), 427–438. https://doi.org/10.1006/nimg.2001.0812

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

Capa, R. L., Duval, C. Z., Blaison, D., & Giersch, A. (2014). Patients with schizophrenia selectively impaired in temporal order judgments. *Schizophrenia Research*, *156*(1), 51–55. https://doi.org/10.1016/j.schres.2014.04.001

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-Pena, M., Holguin, S. R., 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. https://doi.org/10.1111/1469-8986.3660693

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. https://doi.org/10.1111/1469-8986.3660693

Cecere, R., Gross, J., & Thut, G. (2016). Behavioural evidence for separate mechanisms of

audiovisual temporal binding as a function of leading sensory modality. *European Journal of Neuroscience*, *43*(12), 1561–1568. https://doi.org/10.1111/ejn.13242

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being first matters: Topographical representational similarity analysis of erp signals reveals separate networks for audiovisual temporal binding depending on the leading sense. *Journal of Neuroscience*, *37*(21), 5274–5287. https://doi.org/10.1523/JNEUROSCI.2926-16.2017

Chen, S., & Li, X. (2012). Functional magnetic resonance imaging for imaging neural activity in the human brain: The annual progress. *Computational and Mathematical Methods in Medicine*, 2012, 613465. https://doi.org/10.1155/2012/613465

Chen, Y. C., Shore, D. I., Lewis, T. L., & Maurer, D. (2016). The development of the perception of audiovisual simultaneity. *Journal of Experimental Child Psychology*, *146*, 17–33. https://doi.org/10.1016/j.jecp.2016.01.010

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping*, *2*(3), 170–187. https://doi.org/10.1002/hbm.460020306

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). *Identification of Early Visual Evoked Potential Generators by Retinotopic and Topographic Analyses*. *Human Brain Mapping, 2*(3), 170-187. https://doi.org/10.1002/hbm.460020306

Clementz, B. A., Barber, S. K., & Dzau, J. R. (2002). Knowledge of Stimulus Repetition Affects the Magnitude and Spatial Distribution of Low-Frequency Event-Related Brain Potentials. *Audiol Neurootol, 7*(5), 303-314. https://doi.org/10.1159/000064444.

Cong, L. J., Wang, R. J., Yu, C., & Zhang, J. Y. (2016). Perceptual learning of basic visual features remains task specific with Training-Plus-Exposure (TPE) training. *Journal of Vision*, *16*(3), 13–13. https://doi.org/10.1167/16.3.13

Conrey, B., & Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *The Journal of the Acoustical Society of America*, *119*(6), 4065–4073. https://doi.org/10.1121/1.2195091

Cornsweet, T. N. (1962). The staircrase-method in psychophysics. *The American Journal of Psychology*, *75*, 485–491. https://doi.org/10.2307/1419876

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron*, *32*(4), 737–745. https://doi.org/10.1016/S0896-6273(01)00499-8

Curtis, D. R., & Eccles, J. C. (1959). The time courses of excitatory and inhibitory synaptic actions. *The Journal of Physiology*, *145*(3), 529–546. https://doi.org/10.1113/jphysiol.1959.sp006159

Dalton, P. (2000). Psychophysical and Behavioral Characteristics of Olfactory Adaptation. *Chemical Senses*, *25*(4), 487–492. https://doi.org/10.1093/chemse/25.4.487

de Boer-Schellekens, L., Eussen, M., & Vroomen, J. (2012). Diminished sensitivity of audiovisual temporal order in autism spectrum disorder. *Frontiers in Integrative Neuroscience*, *7*, 8. https://doi.org/10.3389/fnint.2013.00008

De Niear, M. A., Koo, B., & Wallace, M. T. (2016). Multisensory perceptual learning is dependent upon task difficulty. *Experimental Brain Research*, *234*(11), 3269–3277. https://doi.org/10.1007/s00221-016-4724-3

De Niear, M. A., Noel, J. P., & Wallace, M. T. (2017). The Impact of Feedback on the Different Time Courses of Multisensory Temporal Recalibration. *Neural Plast, 2017,* https://doi.org/10.1155/2017/3478742

Di Luca, M., Machulla, T. K., & Ernst, M. O. (2009). Recalibration of multisensory

simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12), 7–7. https://doi.org/10.1167/9.12.7

Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*(2), 95–111. https://doi.org/10.1002/hbm.10010

Di Russo, F., & Spinelli, D. (1999). Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vision Research*, *39*(18), 2975–2985. https://doi.org/10.1016/S0042-6989(99)00031-0

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception and Psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/BF03195006

Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological Review*, *122*(2), 232–241. https://doi.org/10.1037/a0038696

Pollack, J. D. (1968). Reaction time to different wavelengths at various luminances. *Perception & Psychophysics, 3,* 17–24. https://doi.org/10.3758/BF03212706

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719–721. https://doi.org/10.1068/p090719

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Research*, *37*(14), 1885–1895. https://doi.org/10.1016/s0042-6989(96)00308-2

Fahle, Manfred., & Poggio, T. (2002). *Perceptual learning*. Massachusetts: MIT Press.

Fahle, Manfred. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*(2), 154–160. https://doi.org/10.1016/j.conb.2005.03.010

Feldman, J. I., Dunham, K., Cassidy, M., Wallace, M. T., Liu, Y., & Woynaroski, T. G. (2018). Audiovisual multisensory integration in individuals with autism spectrum disorder: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *95*, 220–234. https://doi.org/10.1016/j.neubiorev.2018.09.020

Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., … Fernández, G. (2004). Neural bases of cognitive ERPs: More than phase reset. *Journal of Cognitive Neuroscience*, *16*(9), 1595–1604. https://doi.org/10.1162/0898929042568514

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

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

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44. https://doi.org/10.1038/287043a0

Forte, A. E., Etard, O., & Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *ELife*, *6*. https://doi.org/10.7554/eLife.27203

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research*, *203*(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism

spectrum disorders. *Experimental Brain Research*, *203*(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778. https://doi.org/10.1038/nn1268

García-Pérez, M. A. (2014). Adaptive psychophysical methods for nonmonotonic psychometric functions. *Attention, Perception, and Psychophysics*, *76*(2), 621–641. https://doi.org/10.3758/s13414-013-0574-2

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

García-Pérez, M. A., & Alcalá-Quintana, R. (2015). Visual and Auditory Components in the Perception of Asynchronous Audiovisual Speech. *I-Perception*, *6*(6), 204166951561573. https://doi.org/10.1177/2041669515615735

Ghazanfar, A. A., Chandrasekaran, C., & Logothetis, N. K. (2008). Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *Journal of Neuroscience*, *28*(17), 4457–4469. https://doi.org/10.1523/JNEUROSCI.0541-08.2008

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490. https://doi.org/10.1162/089892999563544

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*, 350–363. https://doi.org/10.1038/nrn3476

Gilliatt, R. W., Melville, I. D., Velate, A. S., & Willison, R. G. (1965). A study of normal nerve action potentials using an averaging technique (barrier grid storage tube). *J. Neurol. Neurosurg. Psychiat, 28*(3), 191-200. https://doi.org/10.1136/jnnp.28.3.191

Gohil, K., Hahne, A., & Beste, C. (2016). Improvements of sensorimotor processes during action cascading associated with changes in sensory processing architecture-insights from sensory deprivation. *Scientific Reports*, *6*(1), 1–10. https://doi.org/10.1038/srep28259

Gold, J. I., & Watanabe, T. (2010). Perceptual learning. *Current Biology*, *20*(2), 93-188. https://doi.org/10.1016/j.cub.2009.10.066

Gondan, M., & Heckel, A. (2008). Testing the race inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology*, *52*(5), 322–325. https://doi.org/10.1016/j.jmp.2008.08.002

Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, and Psychophysics*, *78*(3), 723–735. https://doi.org/10.3758/s13414-015-1018-y

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception and Psychophysics*, *67*, 713–726. https://doi.org/10.3758/BF03193527

Gould, R. L., Brown, R. G., Owen, A. M., Ffytche, D. H., & Howard, R. J. (2003). fMRI BOLD response to increasing task difficulty during successful paired associates learning. *NeuroImage*, *20*(2), 1006–1019. https://doi.org/10.1016/S1053-8119(03)00365-3

Green, C. S., Kattner, F., Siegel, M. H., Kersten, D., & Schrater, P. R. (2015). Differences in perceptual learning transfer as a function of training task. *Journal of Vision*, *15*(10), 5–5. https://doi.org/10.1167/15.10.5

Grice, G. R. (1972). Application of a variable criterion model to auditory reaction time as a function of the type of catch trial. *Perception & Psychophysics*, *12*(1), 103–107. https://doi.org/10.3758/BF03212853

Habets, B., Bruns, P., & Röder, B. (2017). Experience with crossmodal statistics reduces the sensitivity for audio-visual temporal asynchrony. *Scientific Reports, 7,* 3–9. https://doi.org/10.1038/s41598-017-01252-y

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Experimental Brain Research*, *166*(3–4), 474–480. https://doi.org/10.1007/s00221-005-2387-6

Hallpike, C. S., & Smith, A. F. R. (1934). The Helmholtz resonance theory of hearing. *Nature*, *133*, 614. https://doi.org/10.1038/133614a0

Harrar, V., Harris, L. R., & Spence, C. (2017). Multisensory integration is independent of perceived simultaneity. *Experimental Brain Research*, *235*(3), 763–775. https://doi.org/10.1007/s00221-016-4822-2

Harris, L., Harrar, V., Jaekl, P., & Kopinska, A. (2010). Mechanisms of simultaneity constancy. In R. Nijhawan (Ed.). *Space and Time in Perception and Action* (pp. 232–253). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750540.015

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

Harrison, Neil R., Witheridge, S., Makin, A., Wuerger, S. M., Pegna, A. J., & Meyer, G. F. (2015). The effects of stereo disparity on the behavioural and electrophysiological correlates of perception of audio-visual motion in depth. *Neuropsychologia*, *78*, 51–

62. https://doi.org/10.1016/j.neuropsychologia.2015.09.023

Harrison, Neil R., & Ziessler, M. (2016). Effect Anticipation Affects Perceptual, Cognitive, and Motor Phases of Response Preparation: Evidence from an Event-Related Potential (ERP) Study. *Frontiers in Human Neuroscience*, *10*, 5. https://doi.org/10.3389/fnhum.2016.00005

Hassenstein, B., & Reichardt, W. (1958). Functional structure of a mechanism of perception of optical movement. *1er Congres International de Cybernetique,* 797–801.

He, J. (2003). Corticofugal modulation of the auditory thalamus. *Experimental Brain Research*, *153*(4), 579–590. https://doi.org/10.1007/s00221-003-1680-5

Heeger, D. (2006). Perception Lecture Notes: Psychophysics. Retrieved March 7, 2021, from http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/psychophysics/psychophysics.html

Henderson, L. (1970). Simple reaction time, statistical decision theory, and the speed-slowness tradeoff. *Psychonomic Science*, *21*(6), 323–324. https://doi.org/10.3758/BF03335803

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, *7*(13). https://doi.org/10.1167/7.13.5

Herzog, M. H., Ewald, K. R. F., Hermens, F., & Fahle, M. (2006). Reverse feedback induces position and orientation specific changes. *Vision Research*, *46*(22), 3761–3770. https://doi.org/10.1016/j.visres.2006.04.024

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141. https://doi.org/10.1016/S0042-6989(97)00043-6

Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, *39*(25), 4232–4243. https://doi.org/10.1016/S0042-6989(99)00138-8

Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, *15*(5), 688–696. https://doi.org/10.1111/j.1467-7687.2012.01171.x

Horsfall, R., Wuerger, S., & Meyer, G. (2021). Visual intensity-dependent response latencies predict perceived audio–visual simultaneity. *Journal of Mathematical Psychology*, *100*, 102471. https://doi.org/10.1016/j.jmp.2020.102471

Hramov, A. E., Maksimenko, V. A., Pchelintseva, S. V., Runnova, A. E., Grubov, V. V., Musatov, V. Y., … Pisarchik, A. N. (2017). Classifying the Perceptual Interpretations of a Bistable Image Using EEG and Artificial Neural Networks. *Frontiers in Neuroscience*, *11*, 674. https://doi.org/10.3389/fnins.2017.00674

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Hülsdünker, T., Ostermann, M., & Mierau, A. (2019). The Speed of Neural Visual Motion Perception and Processing Determines the Visuomotor Reaction Time of Young Elite Table Tennis Athletes. *Frontiers in Behavioral Neuroscience*, *13*, 165. https://doi.org/10.3389/fnbeh.2019.00165

Hunter, M., Godde, B., & Olk, B. (2017). Effects of absolute luminance and luminance contrast on visual discrimination in low mesopic environments. *Attention, Perception, and Psychophysics*, *79*(1), 243–252. https://doi.org/10.3758/s13414-016-1219-z

Husain, M., Shapiro, K., Martin, J., & Kennard, C. (1997). Abnormal temporal dynamics of

visual attention in spatial neglect patients. *Nature*, *385*(6612), 154–156. https://doi.org/10.1038/385154a0

Jaśkowski, P. (1991). Two-stage model for order discrimination. *Perception & Psychophysics*, *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.

Jaskowski, Piotr, Rybarczyk, K., & Jaroszyk, F. (1994). The relationship between latency of auditory evoked potentials, simple reaction time, and stimulus intensity. *Psychological Research*, *56*(2), 59–65. https://doi.org/10.1007/BF00419712

Jeter, P. E., Dosher, B. A., Liu, S. H., & Lu, Z. L. (2010). Specificity of perceptual learning increases with increased training. *Vision Research*, *50*(19), 1928–1940. https://doi.org/10.1016/j.visres.2010.06.016

Jicol, C., Proulx, M. J., Pollick, F. E., & Petrini, K. (2018). Long-term music training modulates the recalibration of audiovisual simultaneity. *Experimental Brain Research*, *236*(7), 1869–1880. https://doi.org/10.1007/s00221-018-5269-4

Johnen, A. K., & Harrison, N. R. (2020). Level of uncertainty about the affective nature of a pictorial stimulus influences anticipatory neural processes: An event-related potential (ERP) study. *Neuropsychologia*, *146*. https://doi.org/10.1016/j.neuropsychologia.2020.107525

Ju, A., Orchard-Mills, E., Van Der Burg, E., & Alais, D. (2019). Rapid Audiovisual Temporal Recalibration Generalises Across Spatial Location. *Multisensory Research*, *32*(3), 215–

234. https://doi.org/10.1163/22134808-20191176

Junghofer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, *37*(4), 523–532. https://doi.org/10.1111/1469-8986.3740523

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11793–11799. https://doi.org/10.1073/pnas.97.22.11793

Kanai, R., Dalmaijer, E. S., Sherman, M. T., Kawakita, G., & Paffen, C. L. E. (2017). Larger Stimuli Require Longer Processing Time for Perception. *Perception*, *46*(5), 605–623. https://doi.org/10.1177/0301006617695573

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(11), 4966–4970. https://doi.org/10.1073/pnas.88.11.4966

Kawakami, S., Uono, S., Otsuka, S., Yoshimura, S., Zhao, S., & Toichi, M. (2020). Atypical Multisensory Integration and the Temporal Binding Window in Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *50*(11), 3944–3956. https://doi.org/10.1007/s10803-020-04452-0

Keane, B., Bland, N. S., Matthews, N., Carroll, T. J., & Wallis, G. (2020). Rapid recalibration of temporal order judgements: Response bias accounts for contradictory results. *European Journal of Neuroscience*, *51*(7), 1697–1710. https://doi.org/10.1111/ejn.14551

Keetels, M., & Vroomen, J. (2011). Perception of synchrony between the senses. In M. M. Murray & M. T. Wallace (Eds.). *The Neural Bases of Multisensory Processes* (pp. 147–

177). Florida: Taylor & Francis. https://doi.org/10.1201/b11092-12

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84. https://doi.org/10.1016/j.plrev.2008.12.001

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Experimental Brain Research*, *60*(3), 492–500. https://doi.org/10.1007/BF00236934

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, *31*(7), 1003–1016. https://doi.org/10.1016/j.neubiorev.2007.03.005

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research*, *18*(1), 65–75. https://doi.org/10.1016/j.cogbrainres.2003.09.004

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

Kohlrausch, A., van Eijk, R., Juola, J. F., Brandt, I., & van de Par, S. (2013). Apparent causality affects perceived simultaneity. *Attention, Perception, and Psychophysics*, *75*(7), 1366–1373. https://doi.org/10.3758/s13414-013-0531-0

Kojo, I., & Pertovaara, A. (1987). The effects of stimulus area and adaptation temperature on warm and heat pain thresholds in man. *International Journal of Neuroscience*,

*32*(3–4), 875–880. https://doi.org/10.3109/00207458709043342

Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception*, *33*(9), 1049–1060. https://doi.org/10.1068/p5169

Kopp, F., & Dietrich, C. (2013). Neural Dynamics of Audiovisual Synchrony and Asynchrony Perception in 6-Month-Old Infants. *Frontiers in Psychology*, *4*, 2. https://doi.org/10.3389/fpsyg.2013.00002

Kovarski, K., Malvy, J., Khanna, R. K., Arsène, S., Batty, M., & Latinus, M. (2019). Reduced visual evoked potential amplitude in autism spectrum disorder, a variability effect? *Translational Psychiatry*, *9*(1), 1–9. https://doi.org/10.1038/s41398-019-0672-6

Krueger Fister, J., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*, *88*, 92–100. https://doi.org/10.1016/j.neuropsychologia.2016.02.016

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered Auditory and Multisensory Temporal Processing in Autism Spectrum Disorders. *Frontiers in Integrative Neuroscience*, *4*, 129. https://doi.org/10.3389/fnint.2010.00129

Lakhani, B., Vette, A. H., Mansfield, A., Miyasike-daSilva, V., & McIlroy, W. E. (2012). Electrophysiological Correlates of Changes in Reaction Time Based on Stimulus Intensity. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036407

Laming, D. (2013). Contrast discrimination by the methods of adjustment and two-alternative forced choice. *Attention, Perception, and Psychophysics*, *75*(8), 1774–1782. https://doi.org/10.3758/s13414-013-0544-8

Lanfer, B., Scherg, M., Dannhauer, M., Knösche, T. R., Burger, M., & Wolters, C. H. (2012). Influences of skull segmentation inaccuracies on EEG source analysis. *NeuroImage*,

*62*(1), 418–431. https://doi.org/10.1016/j.neuroimage.2012.05.006

Le Gros Clark, W. E. (1940). Anatomical basis of colour vision. *Nature*, 146, 558–559. https://doi.org/10.1038/146558b0

Lee, H., & Noppeney, U. (2014). Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00868

Lee, R. W. Y., Jacobson, L. A., Pritchard, A. E., Ryan, M. S., Yu, Q., Denckla, M. B., … Mahone, E. M. (2015). Jitter Reduces Response-Time Variability in ADHD: An Ex-Gaussian Analysis. *Journal of Attention Disorders*, *19*(9), 794–804. https://doi.org/10.1177/1087054712464269

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception and Psychophysics*, *63*(8), 1279–1292. https://doi.org/10.3758/BF03194543

Lengyel, G., & Fiser, J. (2019). The relationship between initial threshold, learning, and generalization in perceptual learning. *Journal of Vision*, *19*(4), 28–28. https://doi.org/10.1167/19.4.28

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. https://doi.org/10.1117/12.912183

Lesicko, A. M. H., & Llano, D. A. (2017, January 1). Impact of peripheral hearing loss on top-down auditory processing. *Hearing Research*, 343, *4–13*. https://doi.org/10.1016/j.heares.2016.05.018

Lewkowicz, D. J., & Flom, R. (2014). The Audiovisual Temporal Binding Window Narrows in Early Childhood. *Child Development*, *85*(2), 685–694. https://doi.org/10.1111/cdev.12142

Li, Q., Xi, Y., Zhang, M., Liu, L., & Tang, X. (2019). Distinct Mechanism of Audiovisual Integration With Informative and Uninformative Sound in a Visual Detection Task: A DCM Study. *Frontiers in Computational Neuroscience*, *13*, 59. https://doi.org/10.3389/fncom.2019.00059

Li, Y., Seger, C., Chen, Q., & Mo, L. (2020). Left Inferior Frontal Gyrus Integrates Multisensory Information in Category Learning. *Cerebral Cortex*, *30*(8), 4410–4423. https://doi.org/10.1093/cercor/bhaa029

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

Lines, C. R., Rugg, M. D., & Milner, A. D. (1984). The effect of stimulus intensity on visual evoked potential estimates of interhemispheric transmission time. *Experimental Brain Research*, *57*(1), 89–98. https://doi.org/10.1007/BF00231135

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

Love, S. A., Petrini, K., Pernet, C. R., Latinus, M., & Pollick, F. E. (2018). Overlapping but Divergent Neural Correlates Underpinning Audiovisual Synchrony and Temporal Order

Judgments. *Frontiers in Human Neuroscience*, *12*, 274. https://doi.org/10.3389/fnhum.2018.00274

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Massachusetts: MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Machulla, T. K., Di Luca, M., Froehlich, E., & Ernst, M. O. (2012). Multisensory simultaneity recalibration: Storage of the aftereffect in the absence of counterevidence. *Experimental Brain Research*, *217*(1), 89–97. https://doi.org/10.1007/s00221-011-2976-5

Machulla, T. K., Luca, M. Di, & Ernst, M. O. (2016). The consistency of crossmodal synchrony perception across the visual, auditory, and tactile senses. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 1026–1038. https://doi.org/10.1037/xhp0000191

Matthews, N., Welch, L., Achtman, R., Fenton, R., & FitzGerald, B. (2016). Simultaneity and temporal order judgments exhibit distinct reaction times and training effects. *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0145926

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience*, *8*(9), 1197–1202. https://doi.org/10.1038/nn1512

McGovern, D. P., Roach, N. W., & Webb, B. S. (2012). Perceptual learning reconfigures the effects of visual adaptation. *Journal of Neuroscience*, *32*(39), 13621–13629. https://doi.org/10.1523/JNEUROSCI.1363-12.2012

McGovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M., & Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience*, *6*, 250. https://doi.org/10.3389/fnagi.2014.00250

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the Multisensory Temporal Window of Integration Results from Changing Task Demands. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071608

Menendez, A., & Lit, A. (1983). Effects of test flash and steady background luminance on simple visual reaction time and perceived simultaneity. *Abstract in: Science*, *24*, unveroeffentlichtes Exemplar. Retrieved from https://pulfrich.siu.edu/Pulfrich_Pages/lit_nonp/1983_Menendez_Lit_Bkgnd.htm

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, 7*(10), 3215-3219. https://doi.org/10.1523/jneurosci.07-10-03215.1987

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Meredith, M. Alex. (2002). On the neuronal basis for multisensory convergence: A brief overview. *Cognitive Brain Research*, *14*(1), 31–40. https://doi.org/10.1016/S0926-6410(02)00059-9

Merzenich, M. M., Kaas, J. H., & Roth, G. L. (1976). Auditory cortex in the grey squirrel: Tonotopic organization and architectonic fields. *The Journal of Comparative*

*Neurology*, *166*(4), 387–401. https://doi.org/10.1002/cne.901660402

Meyer, G. F., Harrison, N. R., & Wuerger, S. M. (2013). The time course of auditory-visual processing of speech and body actions: Evidence for the simultaneous activation of an extended neural network for semantic processing. *Neuropsychologia*, *51*(9), 1716–1725. https://doi.org/10.1016/j.neuropsychologia.2013.05.014

Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, *166*(3–4), 538–547. https://doi.org/10.1007/s00221-005-2394-7

Michael, C. R. (1969). Retinal processing of visual images. *Scientific American*, *220*(5), 105–114. https://doi.org/10.1038/scientificamerican0569-104

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

Mistlin, A. J., & Perrett, D. I. (1990). Visual and somatosensory processing in the macaque temporal cortex: the role of "expectation." *Experimental Brain Research*, *82*(2), 437–450. https://doi.org/10.1007/BF00231263

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

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, *9*(7), 875–877. https://doi.org/10.1038/nn1712

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. *Cerebral Cortex*, *14*(4), 452–465. https://doi.org/10.1093/cercor/bhh007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*, *14*(1), 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

Mullinger, K., & Bowtell, R. (2010). Combining EEG and fMRI. *Methods in Molecular Biology*, *711*, 303–326. https://doi.org/10.1007/978-1-61737-992-5_15

Murray, M. M., & Wallace, M. T. (2011). *The neural bases of multisensory processes.* Florida: Taylor & Francis. https://doi.org/10.1201/9781439812174

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9169–9173. https://doi.org/10.1073/pnas.0810486106

Nelken, I., Ulanovsky, N., Las, L., Bar-Yosef, O., Anderson, M., Chechik, G., … Young, E. D. (2005). Transformation of stimulus representations in the ascending auditory system. In D. Pressnitzer, A. de Cheveigne, S. McAdams & L. Collet, (Eds.). *Auditory Signal Processing* (pp. 264–273). https://doi.org/10.1007/0-387-27045-0_33

Nguyen, T., Kuntzelman, K., & Miskovic, V. (2017). Entrainment of visual steady-state responses is modulated by global spatial statistics. *Journal of Neurophysiology*, *118*(1), 344–352. https://doi.org/10.1152/jn.00129.2017

Nidiffer, A. R., Diederich, A., Ramachandran, R., & Wallace, M. T. (2018). Multisensory perception reflects individual differences in processing temporal correlations. *BioRxiv*,

264457. https://doi.org/10.1101/264457

Niemi, P. (1979). Stimulus intensity effects on auditory and visual reaction processes. *Acta Psychologica*, *43*(4), 299–312. https://doi.org/10.1016/0001-6918(79)90038-6

Nissen, M. J. (1977). Stimulus intensity and information processing. *Perception & Psychophysics*, *22*(4), 338–352. https://doi.org/10.3758/BF03199699

Noel, J. P., De Niear, M. A., Stevenson, R., Alais, D., & Wallace, M. T. (2017). Atypical rapid audio-visual temporal recalibration in autism spectrum disorders. *Autism Research*, *10*(1), 121–129. https://doi.org/10.1002/aur.1633

Noel, J. P., De Niear, M., Van der Burg, E., & Wallace, M. T. (2016). Audiovisual Simultaneity Judgment and Rapid Recalibration throughout the Lifespan. *PLOS ONE*, *11*(8), e0161698. https://doi.org/10.1371/journal.pone.0161698

Noel, J. P., Wallace, M. T., Orchard-Mills, E., Alais, D., & Van Der Burg, E. (2015). True and Perceived Synchrony are Preferentially Associated with Particular Sensory Pairings. *Scientific Reports*, *5,* 17467. https://doi.org/10.1038/srep17467

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735. https://doi.org/10.1038/nn.3248

Pardo-Vazquez, J. L., Padrón, I., Fernández-Rey, J., & Acuña, C. (2014). EEG activity represents the correctness of perceptual decisions trial-by-trial. *Frontiers in Behavioral Neuroscience*, *8*, 105 . https://doi.org/10.3389/fnbeh.2014.00105

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications*, *7*, 1–9. https://doi.org/10.1038/ncomms11543

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology*, *25*(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Peterson, N. N., Schroeder, C. E., & Arezzo, J. C. (1995). Neural generators of early cortical somatosensory evoked potentials in the awake monkey. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials*, *96*(3), 248–260. https://doi.org/10.1016/0168-5597(95)00006-E

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

Pins, D., & Bonnet, C. (1997). Reaction times reveal the contribution of the different receptor components in luminance perception. *Psychonomic Bulletin and Review*, *4*(3), 359–366. https://doi.org/10.3758/BF03210793

Plainis, S., Petratou, D., Giannakopoulou, T., Radhakrishnan, H., Pallikaris, I. G., & Charman, W. N. (2013). Interocular differences in visual latency induced by reduced-aperture monovision. *Ophthalmic and Physiological Optics*, *33*(2), 123–129. https://doi.org/10.1111/opo.12018

Podivinský, F. (1967). Effect of stimulus intensity on the rising phase of the nerve action potential in healthy subjects and in patients with peripheral nerve lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *30*(3), 227–232.

https://doi.org/10.1136/jnnp.30.3.227

Pöppel, E., Schill, K., & von Steinbüchel, N. (1990). Sensory integration within temporally neutral systems states: A hypothesis. *Naturwissenschaften*, *77*(2), 89–91. https://doi.org/10.1007/BF01131783

Powers, A. R., Hevey, M. A., & Wallace, M. T. (2012). Neural correlates of multisensory perceptual learning. *Journal of Neuroscience*, *32*(18), 6263–6274. https://doi.org/10.1523/JNEUROSCI.6138-11.2012

Powers, A. R., Hillock-Dunn, A., & Wallace, M. T. (2016). Generalization of multisensory perceptual learning. *Scientific Reports*, *6*(1), 1–9. https://doi.org/10.1038/srep23374

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

Pratt, H., Martin, W., Schwegler, J., Rosenwasser, R., Rosenberg, S., & Flamm, E. (1992). Temporal correspondence of intracranial, cochlear and scalp-recorded human auditory nerve action potentials. *Electroencephalography and clinical Neurophysiology*, *84*(5), 447-455. https://doi.org/10.1016/0168-5597(92)90032-7.

Preswick, G. (1965). The effect of stimulus intensity on motor latency in the carpal tunnel syndrome. *Psychological Bulletin*, *63*(3), 180–196. https://doi.org/10.1037/h0021700

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590. https://doi.org/10.1111/j.2164-0947.1962.tb01433.x

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

Ran, G., & Chen, X. (2017). The impact of top-down prediction on emotional face processing in social anxiety. *Frontiers in Physiology*, *8*, 1269. https://doi.org/10.3389/fpsyg.2017.01269

Rauschecker, J. P. (2015). Auditory Cortex. In A. W. Toga (Ed.). *Brain Mapping: An Encyclopedic Reference* (Vol. 2, pp. 299–304). Academic press. https://doi.org/10.1016/B978-0-12-397025-1.00226-8

Reale, R. A., & Imig, T. J. (1980). Tonotopic organization in auditory cortex of the cat. *The Journal of Comparative Neurology*, *192*(2), 265–291. https://doi.org/10.1002/cne.901920207

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, *258*(1–2), 89–99. https://doi.org/10.1016/j.heares.2009.04.009

Recanzone, G. H., Guard, D. C., & Phan, M. L. (2000). Frequency and Intensity Response Properties of Single Neurons in the Auditory Cortex of the Behaving Macaque Monkey. *Journal of Neurophysiology, 83*(4), 2315-31. https://doi.org/ 10.1152/jn.2000.83.4.2315

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

Regener, P., Love, S., Petrini, K., & Pollick, F. (2014). Audiovisual processing differences in autism spectrum disorder revealed by a model-based analysis of simultaneity and temporal order judgments. *Journal of Vision*, *14*(10), 429–429. https://doi.org/10.1167/14.10.429

Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P.

(2009). Multisensory integration of sounds and vibrotactile stimuli in processing streams for “what” and “where.” *Journal of Neuroscience*, *29*(35), 10950–10960. https://doi.org/10.1523/JNEUROSCI.0910-09.2009

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., … Vanduffel, W. (2012). Toward a Unified Theory of Visual Area V4. *Neuron*, *74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Rorden, C., Mattingley, J. B., Karnath, H. O., & Driver, J. (1997). Visual extinction and prior entry: Impaired perception of temporal order with intact motion perception after unilateral parietal damage. *Neuropsychologia*, *35*(4), 421–433. https://doi.org/10.1016/S0028-3932(96)00093-0

Roseboom, W. (2019). Serial dependence in timing perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(1), 100–110. https://doi.org/10.1037/xhp0000591

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

Rowland, B. A., & Stein, B. E. (2007). Multisensory integration produces an initial response enhancement. *Frontiers in Integrative Neuroscience*, *1*, 4. https://doi.org/10.3389/neuro.07.004.2007

Rutecki, P. A. (1992). Neuronal excitability: Voltage-dependent currents and synaptic transmission. *Journal of Clinical Neurophysiology*, *9*(2), 195–211. https://doi.org/10.1097/00004691-199204010-00003

Samuelsson, J. G., Khan, S., Sundaram, P., Peled, N., & Hämäläinen, M. S. (2019). Cortical Signal Suppression (CSS) for Detection of Subcortical Activity Using MEG and EEG. *Brain Topography*, *32*(2), 215–228. https://doi.org/10.1007/s10548-018-00694-5

Scarpina, F., Migliorati, D., Marzullo, P., Mauro, A., Scacchi, M., & Costantini, M. (2016).

Altered multisensory temporal integration in obesity. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep28382

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology*, *47*(4), 333–366. https://doi.org/10.1016/S0010-0285(03)00035-5

Schomaker, J. (2009). The Relationship between Response Time and the Strength of Top-Down Attentional Control: An ERP Study. *Journal of European Psychology Students*, *1*(1), 2. https://doi.org/10.5334/jeps.ab

Schröder, A., van Diepen, R., Mazaheri, A., Petropoulos-Petalas, D., Soto de Amesti, V., Vulink, N., … C Lee, A. K. (2014). Diminished N1 auditory evoked potentials to oddball stimuli in misophonia patients. *Frontiers in Behavioural Neuroscience, 8*, 123. https://doi.org/10.3389/fnbeh.2014.00123

Seitz, A. R. (2017). Perceptual learning. *Current Biology*, *27*(13), R631–R636. https://doi.org/10.1016/j.cub.2017.05.053

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075

Setti, A., Burke, K. E., Kenny, R. A., & Newell, F. N. (2011). Is inefficient multisensory processing associated with falls in older people? *Experimental Brain Research*, *209*(3), 375–384. https://doi.org/10.1007/s00221-011-2560-z

Setti, A., Stapleton, J., Leahy, D., Walsh, C., Kenny, R. A., & Newell, F. N. (2014). Improving the efficiency of multisensory integration in older adults: Audio-visual temporal discrimination training reduces susceptibility to the sound-induced flash illusion. *Neuropsychologia*, *61*(1), 259–268. https://doi.org/10.1016/j.neuropsychologia.2014.06.027

Sharma, R., Joshi, S., Singh, K. D., & Kumar, A. (2015). Visual evoked potentials: Normative values and gender differences. *Journal of Clinical and Diagnostic Research*, *9*(7), 12–15. https://doi.org/10.7860/JCDR/2015/12764.6181

Shiga, Y., Yamada, T., Ofuji, A., Fujita, Y., Kawamura, T., Inoue, K., … Yeh, M. H. (2001). Effects of Stimulus Intensity on Latency and Conduction Time of Short-Latency Somatosensory Evoked Potentials. *Clinical EEG and Neuroscience*, *32*(2), 75–81. https://doi.org/10.1177/155005940103200206

Shore, D. I., Gray, K., Spry, E., & Spence, C. (2005). Spatial modulation of tactile temporal-order judgments. *Perception*, *34*(10), 1251–1262. https://doi.org/10.1068/p3313

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

Simon, D. M., Nidiffer, A. R., & Wallace, M. T. (2018). Single Trial Plasticity in Evidence Accumulation Underlies Rapid Recalibration to Asynchronous Audiovisual Speech. *Scientific Reports*, *8*(1), 1–14. https://doi.org/10.1038/s41598-018-30414-9

Simon, D. M., Noel, J. P., & Wallace, M. T. (2017). Event related potentials index rapid recalibration to audiovisual temporal asynchrony. *Frontiers in Integrative Neuroscience*, *11*, 8. https://doi.org/10.3389/fnint.2017.00008

Smith, S. J. M. (2005). EEG in the diagnosis, classification, and management of patients with epilepsy. *Neurology in Practice*, *76*, 2–7. https://doi.org/10.1136/jnnp.2005.069245

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Massachusetts: MIT Press.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–

1973. https://doi.org/10.1162/jocn.2007.19.12.1964

Stelmach, L. B., & Herdman, C. M. (1991). Directed Attention and Perception of Temporal Order. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 539-550. https://doi.org/ 10.1037//0096-1523.17.2.539

Sternberg, S. (2004). Proseminar in Psychological Methods. Retrieved March 7, 2021, from https://www.sas.upenn.edu/~saul/rt.experimentation.pdf

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.

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., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223. https://doi.org/10.1016/j.neuroimage.2008.09.034

Stevenson, R. A., Park, S., Cochran, C., McIntosh, L. G., Noel, J. P., Barense, M. D., … Wallace, M. T. (2017). The associations between multisensory temporal processing and symptoms of schizophrenia. *Schizophrenia Research*, *179*, 97–103. https://doi.org/10.1016/j.schres.2016.09.035

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2018). The cascading influence of multisensory processing on speech perception in autism. *Autism*, *22*(5), 609–624.

https://doi.org/10.1177/1362361317704413

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, *34*(3), 691–697. https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261. https://doi.org/10.1007/s00221-013-3507-3

Stevenson, R. A., Wilson, M. M., Powers, A. R., & Wallace, M. T. (2013). The effects of visual training on multisensory temporal processing. *Experimental Brain Research*, *225*(4), 479–489. https://doi.org/10.1007/s00221-012-3387-y

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

Stockdale, L. A., Morrison, R. G., Kmiecik, M. J., Garbarino, J., & Silton, R. L. (2015). Emotionally anesthetized: Media violence induces neural changes during emotional face processing. *Social Cognitive and Affective Neuroscience*, *10*(10), 1373–1382. https://doi.org/10.1093/scan/nsv025

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215. https://doi.org/10.1121/1.1907309

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. *Behavioral Ecology and Sociobiology, 65*, 13–21 https://doi.org/10.1007/s00265-010-1037-6

Tagliabue, C. F., Veniero, D., Benwell, C. S. Y., Cecere, R., Savazzi, S., & Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-41024-4

Tayeb, Z., Bose, R., Dragomir, A., Osborn, L. E., Thakor, N. V., & Cheng, G. (2020). Decoding of Pain Perception using EEG Signals for a Real-Time Reflex System in Prostheses: A Case Study. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-62525-7

ten Oever, S., Sack, A., Wheat, K. L., Bien, N., & van Atteveldt, N. (2013). Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Frontiers in Psychology*, *4*, 331. https://doi.org/10.3389/fpsyg.2013.00331

Teplan, M. (2002). Fundamentals of EEG Measurement. *Measurement Science Review 2*(2), 1-11.

Thorne, J. D., & Debener, S. (2014). Look now and hear what's coming: On the functional role of cross-modal phase reset. *Hearing Research*, *307*, 144–152. https://doi.org/10.1016/j.heares.2013.07.002

Thurtell, M. J., Bala, E., Yaniglos, S. S., Rucker, J. C., Peachey, N. S., & Leigh, R. J. (2009). Evaluation of optic neuropathy in multiple sclerosis using low-contrast visual evoked potentials. *Neurology*, *73*(22), 1849–1857. https://doi.org/10.1212/WNL.0b013e3181c3fd43

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*(17), 2503–

2522. https://doi.org/10.1016/0042-6989(95)00016-X

Tsujita, M., & Ichikawa, M. (2016). Awareness of Temporal Lag is Necessary for Motor–Visual Temporal Recalibration. *Frontiers in Integrative Neuroscience*, *9*, 64. https://doi.org/10.3389/fnint.2015.00064

Turi, M., Karaminis, T., Pellicano, E., & Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific Reports*, *6*. https://doi.org/10.1038/srep21756

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

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. *Perception & Psychophysics*, *43*(3), 224-239.

Ulrich, Rolf, 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

Ursino, M., Cuppini, C., Magosso, E., Serino, A., & Pellegrino, G. (2009). Multisensory integration in the superior colliculus: A neural network model. *Journal of Computational Neuroscience*, *26*(1), 55–73. https://doi.org/10.1007/s10827-008-

0096-4

Väisänen, O. (2008). Multichannel EEG Methods to Improve the Spatial Resolution of Cortical Potential Distribution and the Signal Quality of Deep Brain Sources. Retrieved March 1, 2021, from https://trepo.tuni.fi/handle/10024/115136

Van der Burg, E., Alais, D., & Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *Journal of Neuroscience*, *33*(37), 14633–14637. https://doi.org/10.1523/JNEUROSCI.1182-13.2013

Van der Burg, E., Alais, D., & Cass, J. (2018). Rapid recalibration to audiovisual asynchrony follows the physical—not the perceived—temporal order. *Attention, Perception, and Psychophysics*, *80*(8), 2060–2068. https://doi.org/10.3758/s13414-018-1540-9

Van Der Burg, E., Alais, D., & Cass, J. (2015). Audiovisual temporal recalibration occurs independently at two different time scales. *Scientific Reports*, *5*(1), 14526. https://doi.org/10.1038/srep14526

Van der Burg, E., & Goodbourn, P. T. (2015). Rapid, generalized adaptation to asynchronous audiovisual speech. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804). https://doi.org/10.1098/rspb.2014.3083

Van der Burg, E., Orchard-Mills, E., & Alais, D. (2014). Rapid temporal recalibration is unique to audiovisual stimuli. *Experimental Brain Research*, *233*(1), 53–59. https://doi.org/10.1007/s00221-014-4085-8

van der Molen, M. W., & Keuss, P. J. G. (1979). The Relationship between Reaction Time and Intensity in Discrete Auditory Tasks. *Quarterly Journal of Experimental Psychology*, *31*(1), 95–102. https://doi.org/10.1080/14640747908400709

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Van der Smagt, M. J. (2016). Audiovisual integration in near and far space: effects of changes in distance and

stimulus effectiveness. *Experimental Brain Research*, *234*(5), 1175–1188. https://doi.org/10.1007/s00221-015-4248-2

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*(6), 955–968. https://doi.org/10.3758/pp.70.6.955

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

Van Essen David C. (1990). Information Processing in the Primate Visual System. Retrieved January 12, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK224275/

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Vassilev, A., Mihaylova, M., & Bonnet, C. (2002). On the delay in processing high spatial frequency visual information: Reaction time and VEP latency study of the effect of local intensity of stimulation. *Vision Research*, *42*(7), 851–864. https://doi.org/10.1016/S0042-6989(01)00300-5

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

Vaughan, H. G., Costa, L. D., & Gilden, L. (1966). The functional relation of visual evoked response and reaction time to stimulus intensity. *Vision Research*, *6*(6), 645–656. https://doi.org/10.1016/0042-6989(66)90076-9

Vroomen, J., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cognitive Brain Research*, *22*(1), 32–35. https://doi.org/10.1016/j.cogbrainres.2004.07.003

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, *91*(3), 484–488. https://doi.org/10.1007/BF00227844

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

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383. https://doi.org/10.1016/j.neuron.2007.10.012

Watson, A. B. (1986). Temporal Sensitivity. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.). *Handbook of Perception and Human Performance* (pp. 6-1 - 6-43). New York: Wiley

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

Weiß, K., & Scharlau, I. (2011). Simultaneity and temporal order perception: Different sides of the same coin? Evidence from a visual prior-entry study. *Quarterly Journal of Experimental Psychology*, *64*(2), 394–416. https://doi.org/10.1080/17470218.2010.495783

Wen, P., Opoku-Baah, C., Park, M., & Blake, R. (2020). Judging relative onsets and offsets of audiovisual events. *Vision*, *4*(1). https://doi.org/10.3390/vision4010017

Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research*, *26*(10), 1677–1691. https://doi.org/10.1016/0042-6989(86)90055-6

Winkler, I., Denham, S., & Escera, C. (2013). Auditory Event-related Potentials. In R . J. Jaeger (Ed.). *Encyclopedia of Computational Neuroscience* (pp. 1–29). New York: Springer. https://doi.org/10.1007/978-1-4614-7320-6_99-1

Wise, A., & Barnett-Cowan, M. (2018). Perceived Simultaneity and Temporal Order of Audiovisual Events Following Concussion. *Frontiers in Human Neuroscience*, *12*, 139. https://doi.org/10.3389/fnhum.2018.00139

Wodka, E. L., Simmonds, D. J., Mahone, E. M., & Mostofsky, S. H. (2009). Moderate variability in stimulus presentation improves motor response control. *Journal of Clinical and Experimental Neuropsychology*, *31*(4), 483–488. https://doi.org/10.1080/13803390802272036

Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., & Krakauer, J. W. (2017). Reaction times can reflect habits rather than computations. *ELife*, *6,* e28075. https://doi.org/10.7554/eLife.28075

Woods, D. L., Alain, C., Covarrubias, D., & Zaidel, O. (1993). Frequency-related differences in the speed of human auditory processing. *Hearing Research*, *66*(1), 46–52. https://doi.org/10.1016/0378-5955(93)90258-3

Xu, J., Sun, X., Zhou, X., Zhang, J., & Yu, L. (2014). The cortical distribution of multisensory neurons was modulated by multisensory experience. *Neuroscience*, *272*, 1–9. https://doi.org/10.1016/j.neuroscience.2014.04.068

Xu, Jinghong, Yu, L., Stanford, T. R., Rowland, B. A., & Stein, B. E. (2015). What does a neuron learn from multisensory experience? *Journal of Neurophysiology*, *113*(3), 883–889. https://doi.org/10.1152/jn.00284.2014

Yarrow, K. (2018). Collecting and Interpreting Judgments about Perceived Simultaneity: A Model-Fitting Tutorial. In A. Vatakis, F. Balcı, M. D. Luca & Á. Correa (Eds.). *Timing and time perception: procedures, measures, & applications* (pp. 295-325). Leiden: Brill.

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

Yarrow, K., Martin, S. E., Di Costa, S., Solomon, J. A., & Arnold, D. H. (2016). A Roving Dual-Presentation Simultaneity-Judgment Task to Estimate the Point of Subjective Simultaneity. *Frontiers in Psychology*, *7*, 416. https://doi.org/10.3389/fpsyg.2016.00416

Yarrow, K., & Roseboom, W. (2017). *Should multisensory temporal acuity be viewed through the window of perceived simultaneity?* https://doi.org/10.31234/osf.io/zpkq7

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

Zerr, M., Freihorst, C., Schütz, H., Sinke, C., Müller, A., Bleich, S., … Szycik, G. R. (2019). Brief Sensory Training Narrows the Temporal Binding Window and Enhances Long-Term Multimodal Speech Perception. *Frontiers in Psychology*, *10*, 2489. https://doi.org/10.3389/fpsyg.2019.02489

Zhou, H., Davidson, M., Kok, P., McCurdy, L. Y., de Lange, F. P., Lau, H., & Sandberg, K. (2020). Spatiotemporal dynamics of brightness coding in human visual cortex revealed

by the temporal context effect. *NeuroImage*, *205*, 116277. https://doi.org/10.1016/j.neuroimage.2019.116277

Zmigrod, L., & Zmigrod, S. (2016). On the Temporal Precision of Thought: Individual Differences in the Multisensory Temporal Binding Window Predict Performance on Verbal and Nonverbal Problem Solving Tasks. *Multisensory Research*, *29*(8), 679–701. https://doi.org/10.1163/22134808-00002532