# University of Bradford eThesis

Hanson, James V. M.

The perceived timing of events across different sensory modalities. A psychophysical investigation of multisensory time perception in humans.

PhD Thesis  2009

University of Bradford, School of Optometry and Vision Science.

ABSTRACT V

ACKNOWLEDGEMENTS VI

CHAPTER 1 1

**1.1 Introduction to the functional anatomy and physiology of the visual, auditory and tactile systems 1**
  1.1.1 The visual system 1
  1.1.2 The auditory system 11
  1.1.3 The tactile system 20

**1.2 Introduction to multisensory processing 30**
  1.2.1 Latency of multisensory brain areas 36
  1.2.2 Temporal factors in multisensory integration 40

CHAPTER 2 47

**2.1 Psychophysical measures of temporal processing 47**
  2.1.1 Reaction time 48
  2.1.2 Temporal resolution 56
  2.1.3 Judgments of temporal order and simultaneity 62
  2.1.4 Perceived duration 76
  2.1.5 Temporal rate perception 83
  2.1.6 Sensorimotor timing 86

CHAPTER 3 93

**3.1 Methodology 93**
  3.1.1 Basic principles of signal detection theory 93
  3.1.2 Forced-choice paradigms 101
  3.1.3 Curve fitting 107

**3.2 Introduction to psychophysical theory and methods 110**
  3.2.1 The method of adjustment 110
  3.2.2 The method of limits 111
  3.2.3 The staircase method 113
  3.2.4 The Method of Constant Stimuli 117

**3.3 Methods of assessing relative timing perception in humans 121**
  3.3.1 Temporal order judgments (TOJs) 121
  3.3.2 Simultaneity judgments (SJ) 123
  3.3.3 Comparison of TOJ and SJ 125
  3.3.4 The method of Libet 128

CHAPTER 4 135

**4.1 An Investigation into the Relationship Between Reaction Time and Temporal Order 135**
4.1.1 Introduction 135

**4.2 Measurement of Simple Reaction Time 138**
4.2.1 Subjects 138
4.2.2 Methods and Stimuli 138
4.2.3 Procedures 141
4.2.4 Results 143
4.2.5 Discussion 144

**4.3 Measurement of crossmodal PSS using a TOJ task 144**
4.3.1 Subjects 144
4.3.2 Methods and Stimuli 145
4.3.3 Procedures 146
4.3.4 Results 147
4.3.5 Discussion 153

**4.4 Measurement of Reaction Times under Conditions of Divided Attention 158**
4.4.1 Subjects 158
4.4.2 Methods and Stimuli 159
4.4.3 Procedures 159
4.4.4 Results 161
4.4.5 Discussion 170

**4.5 Measurement of Choice Reaction Time 174**
4.5.1 Subjects 174
4.5.2 Methods and Stimuli 174
4.5.3 Procedures 175
4.5.4 Results 176
4.5.5 Discussion 180

CHAPTER 5 188

**5.1 The effects of asynchrony adaptation on perceived audiovisual, audiotactile and visuotactile temporal order 188**
5.1.1 Introduction 188

**5.2 Assessment of asynchrony adaptation on perception of crossmodal temporal order 196**
5.2.1 Subjects 196
5.2.2 Methods & stimuli 196
5.2.3 Procedures 197
5.2.4 Results 200
5.2.5 Discussion 205

**5.3 Perception of temporal order following adaptation to asynchronous stimulus pairs evaluated with a staircase paradigm 207**
5.3.1 Methods and stimuli 207
5.3.2 Procedures 207

5.3.3 Results 208
5.3.4 Discussion 209

5.4 Analysis of observer sensitivity and transitivity of temporal order judgments 223
5.4.1 Introduction 223
5.4.2 Methods & stimuli 225
5.4.3 Results 225
5.4.4 Discussion 229

CHAPTER 6 233

6.1 Adaptation to spatially disparate asynchronous audiovisual stimulus pairs 233
6.1.1 Introduction 233

6.2 The Head-Related Transfer Function (HRTF) 236
6.2.1 Measurement of HRTFs 239

6.3 Psychophysical validation of HRTFs 241
6.3.1 Subjects 242
6.3.2 Methods and stimuli 242
6.3.3 Procedures 242
6.3.4 Results 243
6.3.5 Discussion 245

6.4 Adaptation to Spatially Disparate Auditory-Visual Lags 245
6.4.1 Subjects 245
6.4.2 Methods and Stimuli 245
6.4.3 Procedures 246
6.4.4 Results 250
6.4.5 Discussion 253

6.5 Influence of attention on asynchrony adaptation and temporal order judgments 259
6.5.1 Introduction 259
6.5.2 Subjects 263
6.5.3 Methods and stimuli 264
6.5.4 Procedures 264
6.5.5 Results 267
6.5.6 Discussion 270

CHAPTER 7 275

7.1 Effects of sensory modality and causality on sensorimotor temporal order judgments 275
7.1.1 Introduction 275

7.2 Investigation of illusory reversal of perceived sensorimotor temporal order in the visuo-motor, auditory-motor and tactile-motor sensorimotor pairings 284

7.2.1 Observers 285
7.2.2 Methods and stimuli 285
7.2.3 Procedures 285
7.2.4 Results 287
7.2.5 Discussion 290

7.3 Temporal tuning of illusory reversal of perceived sensorimotor temporal order 291
7.3.1 Methods and stimuli 292
7.3.2 Procedures 292
7.3.3 Results 293
7.3.4 Discussion 295

7.4 Transfer across modalities of illusory reversal of perceived sensorimotor temporal order 297
7.4.1 Methods and stimuli 298
7.4.2 Procedures 298
7.4.3 Results 299
7.4.4 Discussion 303

7.5 A Bayesian explanation of illusory reversal of perceived temporal order of motor actions and their sensory consequences 306
7.5.1 Introduction 306
7.5.2 Discussion 310

CONCLUSIONS 317

LIST OF ACRONYMS USED 322

REFERENCES 325

# Abstract

The experiments reported within this thesis use psychophysical techniques to examine the factors which determine perceived multisensory timing in humans. Chapters 1 and 2 describe anatomical and psychophysical features of temporal processing, respectively, whilst Chapter 3 introduces the reader to psychophysical methods. Chapter 4 examines the relationship between two measures of sensory latency, reaction time (RT) and crossmodal temporal order judgment (TOJ). Despite task and attentional manipulations the two measures do not correlate, suggesting that they measure some fundamentally different aspect(s) of temporal perception. Chapter 5 examines the effects of adaptation to asynchronous stimulus pairs on perceived audiovisual (AV), audiotactile (AT) and visuotactile (VT) temporal order. Significant temporal shifts are recorded in all three conditions. Evidence is also presented showing that crossmodal TOJs are intransitive. Chapter 6 shows that concurrent adaptation to two sets of asynchronous AV stimulus pairs causes perceived AV temporal order to recalibrate at two locations simultaneously, and that AV asynchrony adaptation effects are significantly affected by observers' attention during adaptation. Finally, Chapter 7 shows that when observers are accustomed to a physical delay between motor actions and sensory events, an event presented at a reduced delay appears to precede the causative motor action. The data are well-described by a simple model based on a strong prior assumption of physical synchrony between motor actions and their sensory consequences.

# Acknowledgements

Over the last three years, I have been fortunate to receive help and support from many people whilst undertaking the research described within this work. I would like to take the opportunity to place on record my gratitude to these individuals.

First and foremost, I would like to express my profound thanks to my supervisors, Professor David Whitaker and Dr James Heron. Over the course of my PhD, I have benefited greatly from their knowledge and experience of this area of research, and of psychophysics generally. I am grateful to them for agreeing to supervise my PhD, and for providing guidance and encouragement throughout in a friendly and supportive manner. I am also grateful to Dr Neil Roach and Professor Paul McGraw from the University of Nottingham; their technical skills and advice have been invaluable over this period. From within the School of Optometry and Vision Science at Bradford, I have also benefited from the moral support and assistance provided by Chris Davey, Mark Burton and Chara Vakrou.

On a personal level, I thank my parents Michael and Valerie for their belief and support over the last 33 years. Finally, I would never have begun a PhD without the encouragement of my girlfriend Isabelle; her optimism, patience and unfailing good humour have been invaluable during the times when these qualities have eluded me.

# Chapter 1

## 1.1 Introduction to the functional anatomy and physiology of the visual, auditory and tactile systems

In this thesis, I investigate the perceived timing of events in the visual, auditory and tactile modalities. To appreciate why perceived event timing may differ between the modalities, an understanding of the functional anatomy of these sensory systems and the physiological factors which influence perceptual latency is important. A brief summary is presented below.

### 1.1.1 The visual system

An eye may be thought of as an optical system that focuses light on the retina (the first stage of neural processing in the visual system). Light rays from an object enter the eye and pass through the cornea, aqueous humour, crystalline lens and vitreous humour before finally arriving at the retina, which covers the posterior surface of the internal eye. In passing through these structures and being brought into focus upon the retina, the light is said to be *refracted*. The gross anatomy of the human eye can be seen in Figure 1.1.

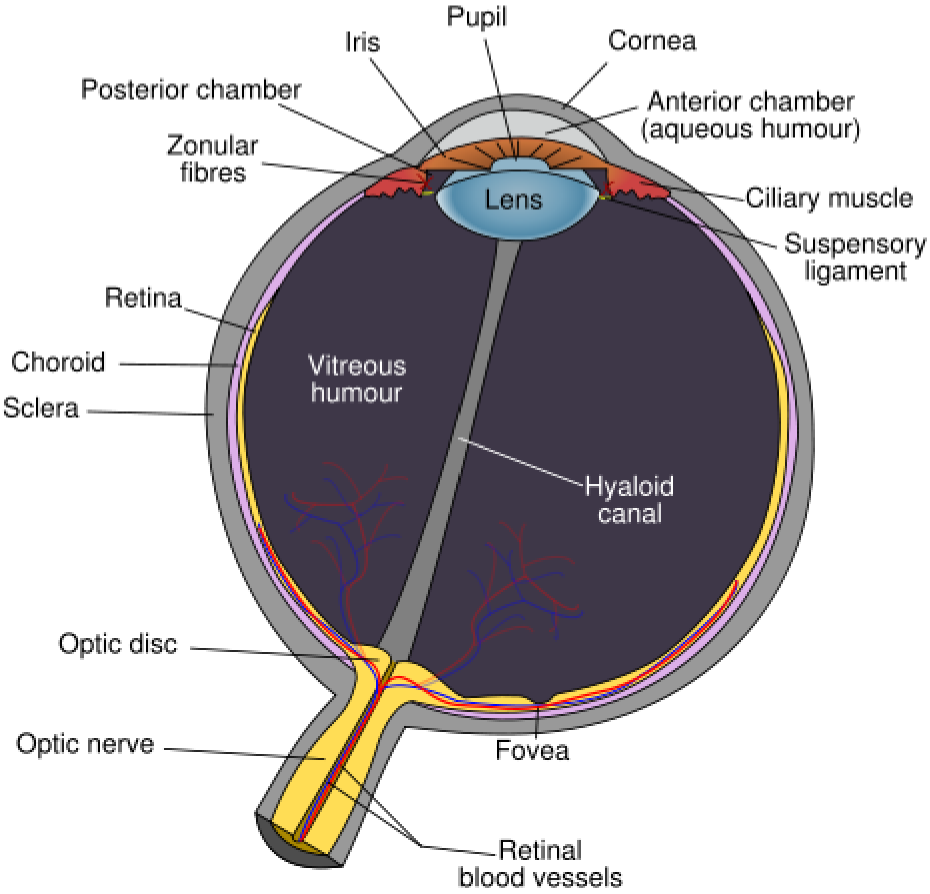


**Figure 1.1.** Diagram showing the structure of the human eye. Light rays are refracted via the cornea, aqueous humour, crystalline lens and vitreous humour before reaching the retina. Taken from www.en.wikipedia.org/wiki/eye.

The retina is the location at which light is converted, or transduced, into neural signals. It contains five types of neuron: photoreceptors, bipolar cells, ganglion cells, horizontal cells, and amacrine cells. These neurones, together with the remaining retinal tissues, are traditionally divided into around eight layers (note that some sources include both an inner and outer limiting membrane in this figure, making a total of ten layers (Snell & Lemp, 1998)). An illustration of the cells and layers of the retina can be seen as Figure 1.2.

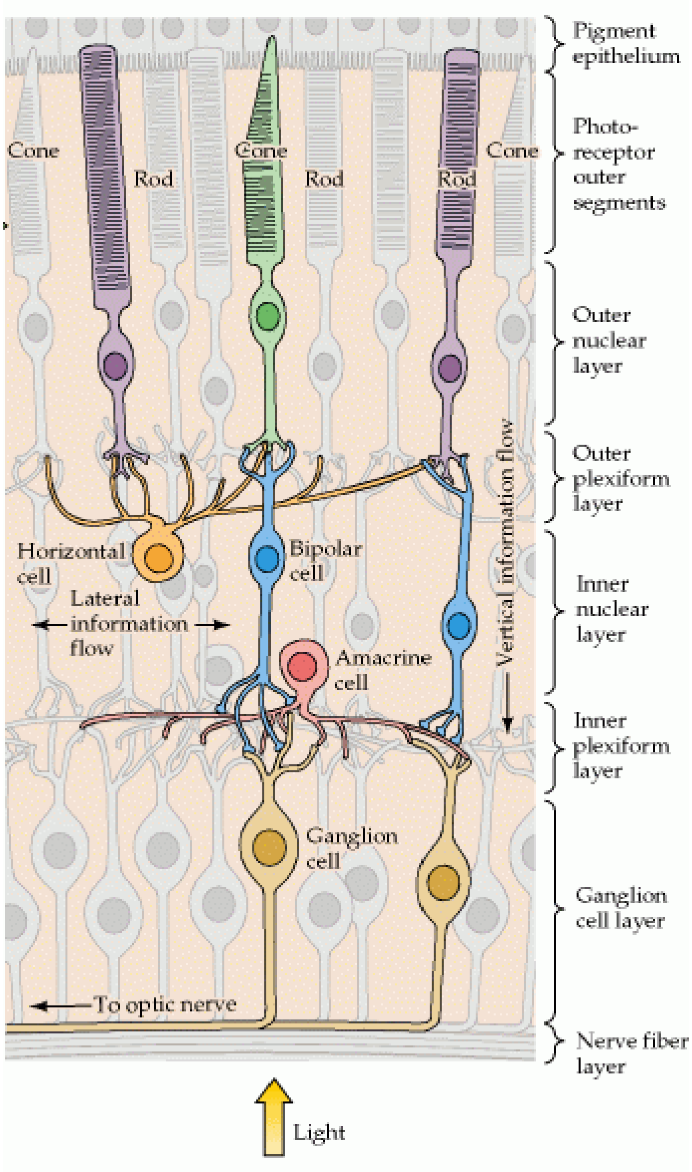


**Figure 1.2.** Diagram showing the different cells and layers of the human retina, including both types of photoreceptors (rods and cones). Note that to reach the photoreceptors and be transduced into a neural signal, incident light first has to pass through the other cells and their axons. The vertical and lateral flow of information is indicated on the diagram (see text for details). Taken from www.ncbi.nlm.nih.gov /books/bv.fcgi?rid=neurosci.figgrp .740

Light hitting the photoreceptors causes neural signals to be passed to the bipolar cells and then to the ganglion cells, the axons of which collect at the optic disc and exit the eye as the optic nerve. As can be seen from inspection of Figure 1.2, this processing route can be thought of as a vertical pathway. In contrast, horizontal and amacrine cells form a lateral processing pathway. Horizontal cells form a connection between photoreceptors and bipolar cells, and are known to sharpen contrast and increase spatial resolution (Snell &

Lemp, 1998). Amacrine cells connect with bipolar cells, ganglion cells and other amacrine cells; there are believed to be between 20-40 different types, but their precise functions are not known (Oyster, 1999; Rodieck, 1998).

As can be seen from Figure 1.2, two types of photoreceptors are known to exist in humans. Known as rods and cones, each photoreceptor type has distinct anatomical and functional features. For this reason, the human retina is often referred to as duplex in nature (Sekuler & Blake, 1994). Rods are much more numerous in number (90-125million vs. 4-8 million), and possess a long, thin, and straight outer segment. They are optimally sensitive to light with a wavelength of approximately 500nm, and operate only in dim (scotopic) light conditions. In contrast, cones possess a shorter, cone-shaped outer segment and operate in bright (photopic) light conditions. Three types of cone exist, optimally sensitive to light of 420, 530 and 560nm (S, M, and L cones, respectively. Differences in the neural signals emanating from the S, M, and L cones enable the subjective sensation of colour. Cone distribution is extremely dense at the fovea (the central area of retina associated with highest spatial resolution), but they are poorly represented elsewhere in the retina. In comparison, rods are absent at the fovea and rise rapidly in number moving away from the fovea, with the number decreasing toward the retinal periphery.

As stated above, the nerve fibres of the retinal ganglion cells exit the eye as the optic nerve. The route taken by nerve impulses to the main site of visual processing in primates, primary visual cortex (V1/Brodmann area 17/striate cortex/occipital cortex), is commonly known as the visual pathway or geniculo-striate pathway. An illustration of this pathway is shown as Figure 1.3. It can be appreciated that the image of an object to the left of an observer will fall on the nasal retina of the left eye, and the temporal retina of the right eye. To combine the images from the two eyes so that neural signals representing a specific area of space reach the same area of visual cortex, the nerve fibres from the nasal retinae of both eyes cross over to the opposite side of the brain at the optic chiasm. This process is termed partial decussation, and has the effect that objects to the left of an observer are processed in the right cerebral hemisphere (and vice versa). After the chiasm, the body of nerve fibres is referred to as the optic tract rather than optic nerve.

The first synapse of the visual pathway is at the lateral geniculate nucleus (LGN) of the thalamus, where the nerve fibres undergo reorganisation by segregation into magnocellular and parvocellular processing streams (Oyster, 1999). Both streams project to different layers within the LGN (Rodieck, 1998). This segregation is functional as well as anatomical; the parvocellular stream has high spatial and low temporal resolution whilst carrying information as to the spectral content of a scene. Conversely, the magnocellular stream has low spatial and high temporal resolution, and is unable to convey spectral content (Frishman, 2005). These functional differences are dictated by the properties of different sub-types of retinal bipolar and ganglion cells which project to the different LGN layers (Frishman, 2005; Oyster, 1999). After the LGN, the nerve fibres are now referred to as the optic radiations until they terminate in primary visual cortex (V1), located in the occipital lobes of the brain. Image processing is initially carried out here before subsequent specialised analysis at areas such as V2/prestriate cortex, V3, V4, and V5/visual area MT (middle temporal).

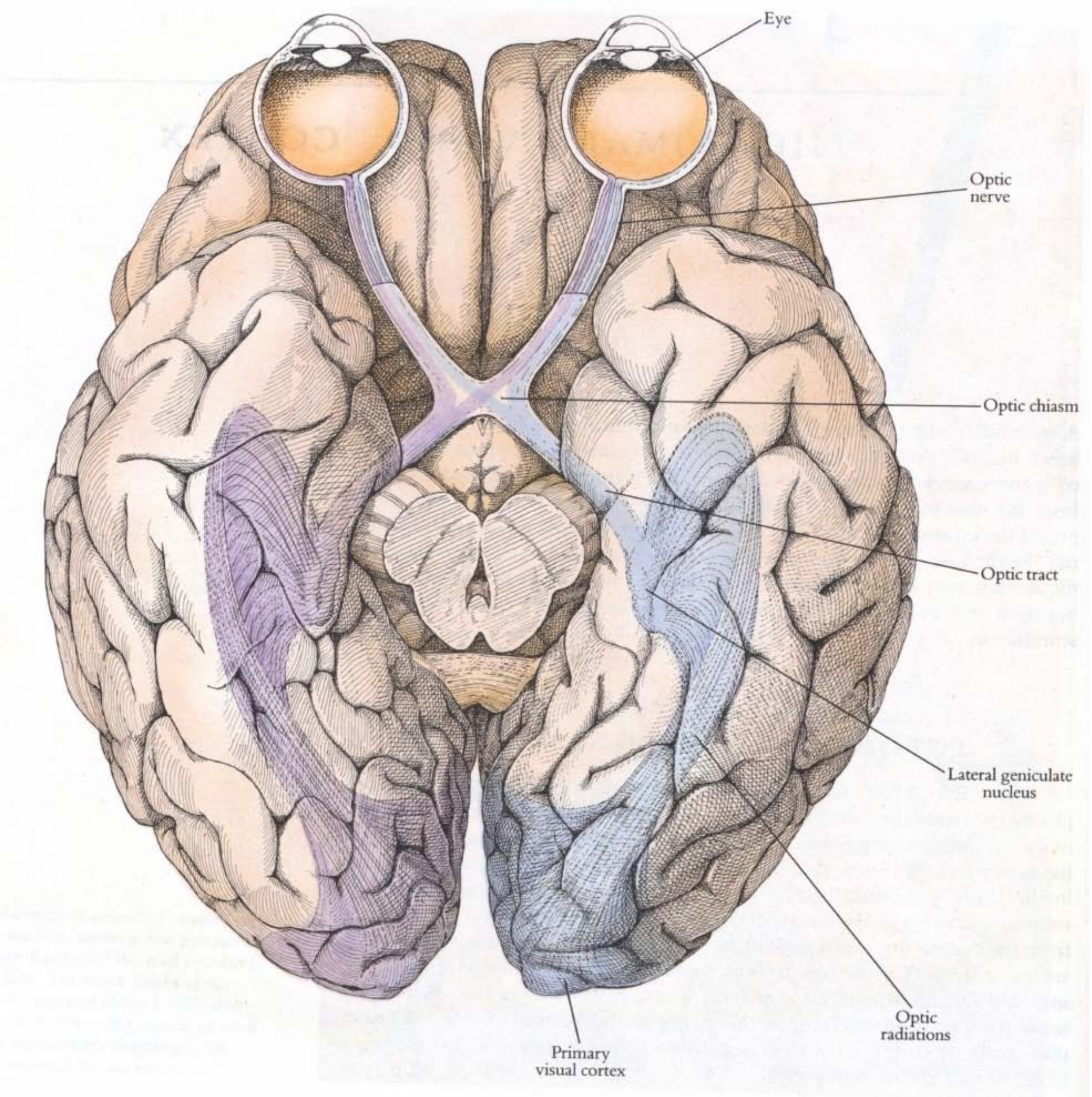


**Figure 1.3.** Diagram showing the structures of the human visual pathway (axial view). Note that primary visual cortex is at the opposite aspect of the brain to the eyes. Taken from http://www.nmr.mgh.harvard.edu/~rhoge/HST583/doc/VisualCortex.jpg.

A small proportion (<10%) of nerve fibres leave the optic tract prior to the LGN and project to the pretectum, superior colliculus, and pregeniculate (Frishman, 2005). The pretectum is involved in the pupillary response to light, whereas the function of the pregeniculate is not currently known (Rodieck, 1998). The superior colliculus is known to direct eye movements to locations and objects of interest (Rodieck, 1998); however, it is also known to integrate information from different sensory modalities (Stein, Wallace, & Stanford, 2005). A more detailed discussion of multisensory integration in the superior colliculus is presented in Sections 1.2.1 and 1.2.2. This second visual pathway is referred to as the subcortical pathway.

A fundamental issue is the time course of the visual process. In other words, if an observer witnesses a visual event, how long will it take for a representation of the event to become available to the perceptual system? This period of perceptual latency may usefully be subdivided into at least three components: the time taken for light to reach the sensory receptor surface, the time taken for transduction of light into a neural signal, and the time taken for the neural signal to travel to the brain. The first of these components can be effectively excluded as a cause of any *significant* delay, as the velocity of light is over 186,000 miles per second in air (slightly slower within the eye).

In order to assess any effect of transduction speed on visual perceptual latency, an understanding of the stages of the transduction process at the retina is essential. Photopigment molecules are contained within the outer segments of rods and cones, and consist of a protein called an opsin combined with a chromophore called 11-*cis* retinaldehyde (Fain, 2003). The opsins differ between the four types of photoreceptor (rods, S-, M- and L-cones), and it is these proteins which give the photopigments their sensitivity to different wavelengths of visible light (Frishman, 2005). Absorption of light by photopigment molecules causes a chemical process called isomerisation, which ultimately results in closure of ion channels in the outer segment (Pasternak, Bisley, & Calkins, 2003). Closure of the ion channels reduces the concentration of calcium ions inside the photoreceptor, causing hyperpolarisation of the photoreceptor. This in turn reduces the concentration of the neurotransmitter glutamate at the photoreceptor synaptic terminal (Wolfe et al., 2006), which signals to the connecting bipolar cell that the photoreceptor has absorbed light – in other words, it represents the end result of the photoreceptors' response to light. The entire process takes a *minimum* of 7msec for both rods and cones following bright light stimulation (Cobbs & Pugh, 1987; Hestrin & Korenbrot, 1990).

However, transduction of light at the photoreceptors is only the first step in visual perception; the neural signal has to travel through the bipolar cells to the ganglion cells, interacting with horizontal and amacrine cells *en route*, before joining the visual pathway proper. Even where the retinal signal path is at its most direct (single foveal cone → two midget bipolar cells → single ganglion cell), three cell types and two synaptic stages are needed to send a visual signal to the brain. Therefore, retinal transduction is effectively a complex biochemical cascade.

The physical length of the visual pathway from retina to brain also affects neural latency, as the neural signal initiated by the photoreceptors has to arrive at cortical areas in order to reach consciousness. Inspection of Figure 1.3 reveals that the retina is at the opposite aspect of the brain from V1 (Woolsey, Hanaway, & Gado, 2003). The time taken for the neural impulse to reach sensory cortex will be determined by this distance and the speed at which the neural impulse travels within the brain.

The speed at which a nerve impulse travels towards V1 (or the speed of any other nerve impulse within the body) can be quantified by the Nerve Conduction Velocity (NCV). NCVs vary according to the diameter and degree of myelination of nerve fibres, with typical velocities for myelinated nerve fibres ranging between 18-120m/sec (Brodal, 2004). Maximal conduction velocity in sensory nerves is usually considered to be in the region of 40-70m/sec (Aminoff, 2003). Note that both of these values relate to peripheral nerves, i.e. those not within the brain or central nervous system.

Within the brain, animal data indicate that axonal diameter and degree of myelination (and, consequently, NCV) is not uniform throughout the sensory pathways (Baker & Stryker, 1990; Salami, Itami, Tsumoto, & Kimura, 2003). The question of NCV in the human visual pathway is one that has received very little research attention; however, two studies have suggested values of 2m/sec (Reed & Jensen, 1992; Reed, Vernon, & Johnson, 2004). In these studies, the length of the visual pathway was not measured directly (e.g., by the use of Magnetic Resonance Imaging (MRI)), but inferred from head length. Head length was divided by the latency of Visual-Evoked Potentials (VEP) in the brain to estimate NCV. Such a method is likely to significantly underestimate NCV as it does not take into account the curvature of the optic radiations, as evident from Figure 1.3 (Johnson, Reed, & Vernon, 2005; Saint-Amour, Saron, Schroeder, & Foxe, 2005). It would also fail to take into account synaptic time at the LGN (approximately 7-10msec in macaque (Maunsell & Gibson, 1992)) and the latency of the retinal response to light (Saint-Amour et al., 2005). Indeed, retinal latency (rather than nerve conduction time) is considered the primary cause of the delay in a visual event becoming available to the perceptual system (Bolz, Rosner, & Wassle, 1982). The value of 2m/sec is therefore extremely unlikely to reflect the true NCV in the human brain. In any case, given the non-uniformity of within-brain NCV in animals, as discussed above, it is debatable how useful NCV in isolation is as a measure of sensory latency. More practical would be a measure of the time taken for activity to be recorded in cortical areas following visual stimulation.

Single-unit recording has demonstrated that a visual event causes activation in macaque V1 an average of 66msec later (range 34-97msec) (Schmolesky et al., 1998), which is comparable to earlier work suggesting a figure of 30-50msec (Maunsell & Gibson, 1992). Both of these studies found activation in other areas of visual cortex to occur after activation in V1, reflecting the hierarchical nature of the primate visual system (Maunsell & Gibson, 1992; Schmolesky et al., 1998). As macaques are physically smaller than humans, it would be expected that the time taken for neural signals to travel through the visual system would be greater in human than macaque. One group of researchers has suggested that a scaling factor may be employed to more accurately relate monkey sensory latency times to humans, with monkey latency being approximately 3/5 that of humans in the equivalent cortical region (Schroeder, Molholm, Lakatos, Ritter, & Foxe, 2004; Schroeder, Seto, Arezzo, & Garraghty, 1995). Taking this into account, we may therefore assume V1 activation in humans a *minimum* of 50-57msec following presentation of a visual stimulus, with higher cortical areas being activated after this. This value is comparable to that measured in humans using scalp-recorded electrophysiological methods (Clark & Hillyard, 1996; Foxe & Simpson, 2002). Considering all these factors in combination, it appears likely that our visual percept of the environment is delayed by a *minimum* of 50msec. In other words, a visual scene we perceive as being in the present is actually in the recent past (Hallett, 2007).

### 1.1.2 The auditory system

The auditory system is responsible for the processing of sound. Sound is the result of some mechanical disturbance (e.g., motion or vibration of an object), which causes the propagation of a pressure wave (i.e., a sound wave) through air. These sound waves interact with three subdivisions of the human auditory system: the outer ear, middle ear and inner ear. The outer ear consists of the external ear (pinna) and auditory canal; the middle ear consists of the tympanic membrane and ossicles (maleus, incus and stapes); and the inner ear consists of the cochlea and auditory nerve (Sekuler & Blake, 1994). Figure 1.4 shows the anatomical structures of the human auditory system.

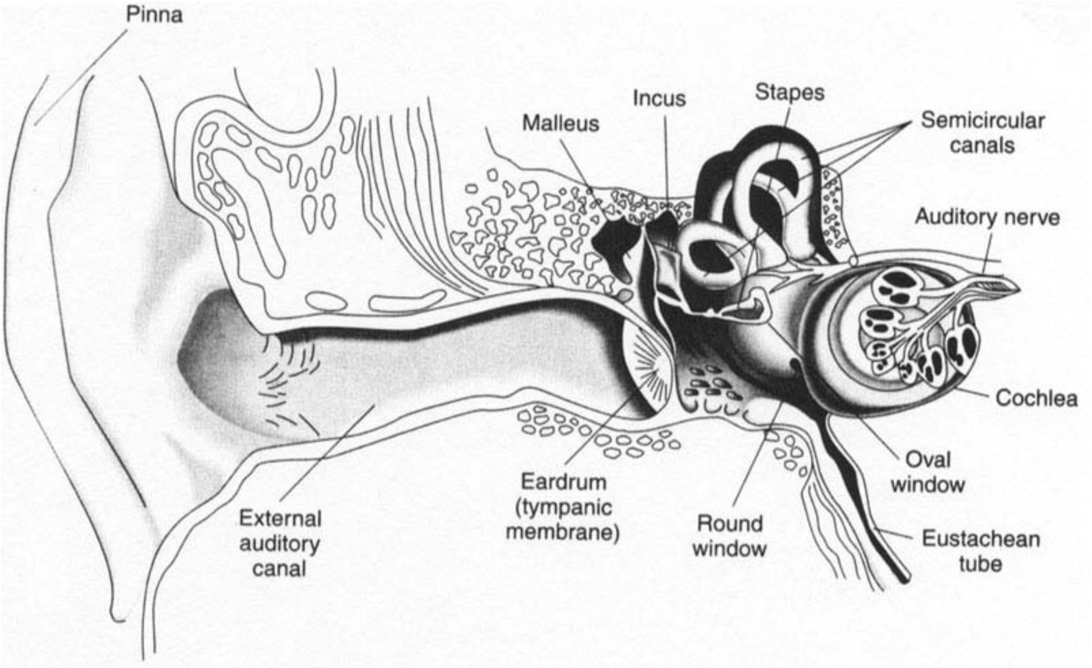


**Figure 1.4.** Diagram showing the anatomy of the human auditory system. Sound waves are filtered by the pinna and enter the auditory canal. When the sound waves meet the tympanic membrane, the vibration of the membrane is transmitted to the oval window of the cochlea by the malleus, incus and stapes (collectively known as the ossicles). From the cochlea, neural signals corresponding to the nature of the incident sound are initiated which travel toward the brain. Taken from www.skidmore.edu/~hfoley/Perc9.htm.

Each of the two pinnae modifies and filters incoming sound waves via the many prominent folds visible on its surface, with similar filtering occurring due to the shape of the head and torso. The results of this filtering process vary according to the spatial origin of the incoming sound waves, and modification at the pinnae is therefore important in the localisation of a sound source (Moore, 2005). Incident sound is then channelled into the auditory canal, an opening into the skull approximately 7mm in diameter, toward the tympanic membrane (eardrum). The auditory canal functions mainly to protect the tympanic membrane and the rest of the auditory structures, but due to its length and shape also amplifies sound frequencies between approximately 2000-6000 Hz (Wolfe et al., 2006).

When sound has travelled through the auditory canal, it reaches the tympanic membrane or eardrum. Pressure from incident sound waves causes the eardrum to vibrate. This vibration is transmitted through the middle ear to the oval window of the cochlea by the ossicles, which also increase the magnitude of the mechanical energy transmitted to the oval window. This magnification is due mainly to the physical difference in size between the tympanic membrane and the oval window, the tympanic membrane being larger by a factor of approximately 18; however, the joints between the ossicles also increase the energy transmitted from one side of the joint to another by mechanical means (Geldard, 1972). The increase in energy transmitted through the middle ear is necessary because the membrane covering the oval window is more mechanically resistant to displacement than air, due to the fluid contained within the cochlea (Moore, 2005).

As the site where incident energy is transduced into neural signals, the cochlea may be considered analogous to the retina of the eye. Structurally, it consists of three parallel, fluid-filled canals - the vestibular canal (scala vestibuli), tympanic canal (scala tympani), and middle canal (scala media), arranged into a spiral. A cross section of the cochlea illustrating these structures is shown in Figure 1.5. Nerve impulses are generated by the inner hair cells of the basilar membrane. Vibration of the basilar membrane causes friction between the hair cells and tectorial membrane; this friction initiates neural discharge of the inner hair cells. The outer hair cells, although not active in transduction, are thought to actively influence sensitivity and frequency tuning via efferent connections from the auditory centres of the brain, primarily the superior olivary complex of the brainstem (Moore, 2005).

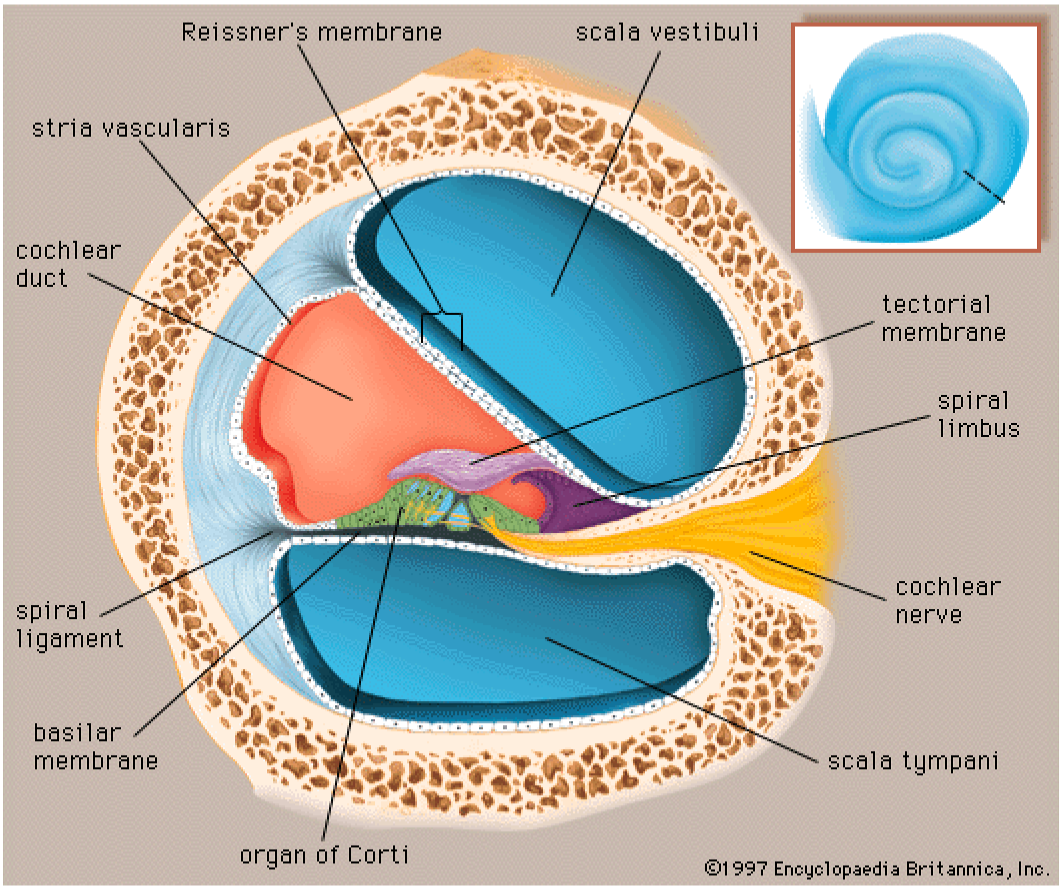


**Figure 1.5.** Cross section of the cochlea showing the three fluid-filled parallel chambers – scala vestibuli, scala tympani, and scala media/cochlear duct. The inner hair cells of the basilar membrane discharge in response to friction with the tectorial membrane. Taken from http://original.britannica.com/eb/art/print?id=534&articleTypeId=0.

Following discharge of the inner hair cells, the neural signals travel along the auditory (vestibulocochlear; cochlear) nerve to the brain. Figure 1.6 shows the main structures and synapses involved in the auditory pathway. The fibres from the auditory nerve first travel to the cochlear nuclei of the midbrain, where they synapse. It is known that frequency processing occurs at this site, with some cochlear nucleus cells coding specific frequencies and inhibiting responses of cells coding adjacent frequencies (Palmer, 1995). The auditory pathway then extends from the ipsilateral cochlear nucleus to *both* superior olive nuclei (located at the pons of the brainstem), a process analogous to the partial decussation of retinal nerve fibres at the optic chiasm. Therefore, this is the first stage in the auditory pathway in which interactions between the signals from both ears (binaural interactions) can be found (Hackney, 1987; Regan, 1989). These binaural interactions convey valuable information about the location of a sound source (Grantham, 1995; Moore, 1987; Sekuler & Blake, 1994).

Upon leaving the superior olive, the emergent tract of nerve fibres becomes the lateral lemniscus. Some of the fibres synapse in the three nuclei of the lateral lemniscus (Hackney, 1987). The nuclei project to the inferior colliculus at the midbrain region of the brainstem, as well as to the contralateral lateral lemniscus (Hackney, 1987). The next structure of the auditory pathway, the inferior colliculus, receives input from both the ipsilateral superior olive and contralateral cochlear nuclei. However, the majority of inferior colliculus input is ultimately from the contralateral ear (Wolfe et al., 2006), and cells here are generally suppressed by ipsilateral input (Regan, 1989). The inferior colliculus is also known to receive somatosensory afferent input, and input from the contralateral inferior colliculus (Hackney, 1987).

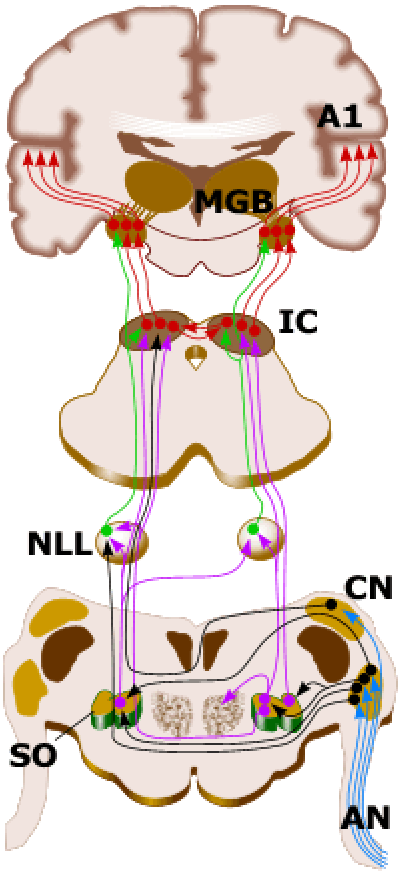


**Figure 1.6.** Diagram showing the main structures of the human ascending auditory pathway. Abbreviations are AN (auditory nerve), CN (cochlear nuclei), SO (superior olive), NLL (nuclei of the lateral lemniscus), IC (inferior colliculus), MGB (medial geniculate body), A1 (primary auditory cortex). Taken from http://www.urmc.rochester.edu/smd/ Nanat/faculty-research/lab-pages/KevinDavis/ passways.gif.

The final structure in the pathway prior to primary auditory cortex is the medial geniculate body of the thalamus. The dorsal division of the medial geniculate body receives visual and somatosensory afferent input in addition to auditory, and thus input from these sensory systems can modulate the responses of a subset of auditory neurons in the medial geniculate (Regan, 1989). From here, the nerve fibres progress to ipsilateral primary auditory cortex (A1), at the superior temporal gyrus of the brain's temporal lobe (Figure 1.7). Cortical processing of auditory events is performed primarily in A1 (Brodmann areas 41 and 42) and associated secondary auditory cortical areas. The spatial extent and function of these areas in humans is not yet fully understood (Ehret & Scheich, 1997), however the analogous areas in non-human primates are referred to as belt and parabelt areas (Hackett & Kaas, 2003). Note that in comparison with the visual system, a large amount of auditory processing is performed prior to A1 (i.e., subcortically), whereas very little visual processing occurs *en route* to V1 (Wolfe et al., 2006). Consequently, the subcortical auditory pathway is extremely complex in comparison with both the visual and tactile modalities (Hackett & Kaas, 2003). Auditory input to the multisensory regions of the superior colliculus is from the inferior colliculus, superior olive, and lateral lemniscus (Edwards, Ginsburgh, Henkel, & Stein, 1979).

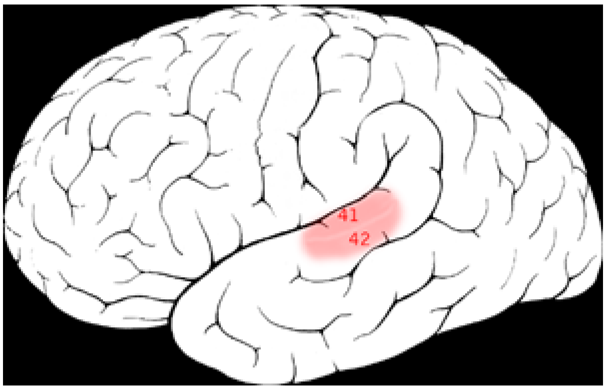


**Figure 1.7.** Sagittal view of the human brain illustrating the location of primary auditory cortex (A1; Brodmann areas 41 & 42). The occipital lobe (containing V1) is at the right of the diagram. Taken from http://en.wikipedia.org/wiki/Image:Brodmann_41_42.png.

As with the visual system, the time taken for an auditory event to be represented in the brain can be divided into the time taken for a signal to reach the sensory apparatus, the time taken by the transduction process, and the time taken for a nerve impulse to travel to auditory cortex. The speed of sound in air is approximately 343m/sec, significantly slower than the speed of light. These differential velocities cause auditory and visual signals arising from distant events to arrive asynchronously at their respective receptor sites, despite synchronous generation at source (Section 2.1.3).

In comparison to the somewhat sluggish speed of sound in air, the transduction of sound waves at the cochlea is an extremely rapid process. The latency of the cochlear hair cell response to stimulation is approximately 40µsec (Corey & Hudspeth, 1979), reflecting the direct mechanical nature of transduction at this site (Fain, 2003). This almost instantaneous process is significantly faster than the complex biochemical cascade that forms the visual transduction process (Section 1.1.1).

In addition to the transduction factors outlined above, the length of the auditory pathway must also be considered. Figure 1.8 illustrates the location of A1 in relation to the cochlea and brainstem. From comparison with Figure 1.3, it can be seen that the auditory pathway is shorter than the visual pathway (Woolsey et al., 2003). Together with the almost instantaneous transduction of auditory events by the inner hair cells of the cochlea (Corey & Hudspeth, 1979), this implies that the auditory component of proximal audiovisual events will be processed before the visual component.

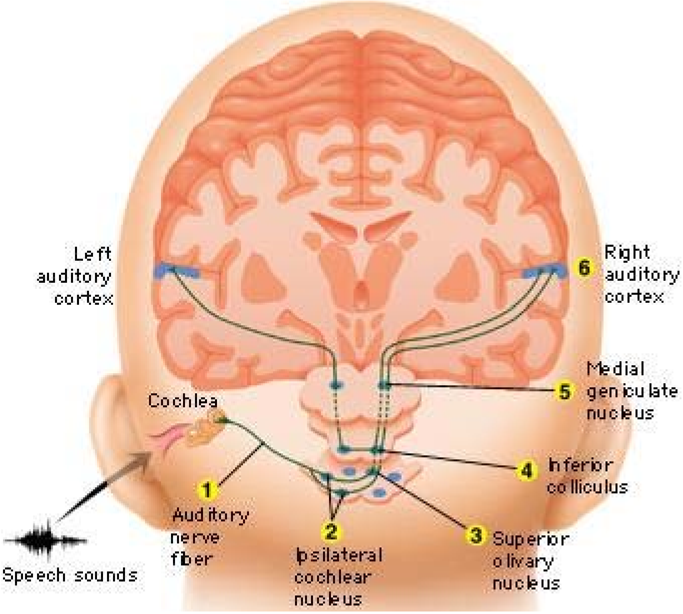


**Figure 1.8.** Simplified illustration of the auditory pathway (coronal view) showing the location of auditory cortex (A1) and its location relative to the cochlea. Taken from http://www.Brainconnection.com/med/medart/l/anat/990705.jpg.

The effect of the extremely rapid transduction of sounds at the cochlea and the short axonal distance of the auditory pathway is that both pure-tone and broadband white noise stimuli first cause activation in macaque A1 approximately 9msec following stimulus onset (Fu et al., 2004; Lakatos et al., 2005). In these studies, the distance of the speaker from the monkey was not recorded, but both data sets were collected in an indoor environment. Assuming that monkey latency is 3/5 that of human latency (Schroeder et al., 2004; Schroeder et al., 1995), this would translate to human A1 first being activated approximately 15-16msec following a proximal auditory event. This value corresponds almost exactly to that measured using magnetoencephalography (MEG) in humans (Gutschalk et al., 1999; Inui, Okamoto, Miki, Gunji, & Kakigi, 2006). Thus, relative to visual events, proximal auditory events first become available to the perceptual system at very short latencies.

It may be appreciated from the preceding discussion that following distant audiovisual events light will reach an observer before sound (due to its significantly greater velocity in air), whereas an event occurring next to an observer will most likely result in sound being processed before light (due to the significantly faster auditory transduction process and physically shorter length of the auditory pathway). At a distance of approximately 10m from an observer, all of these differences negate each other (Poppel, Schill, & von Steinbuchel, 1990). The effect of this is that the simultaneous auditory and visual components of an event 10m from an observer should arrive at their respective primary sensory cortices approximately simultaneously; hence, this distance has been termed the ‘horizon of simultaneity’ (Poppel et al., 1990).

### 1.1.3 The tactile system

What is often referred to as the tactile system (the system which mediates the sense of ‘touch’) is actually a division of the somatosensory system. In addition to touch, sensations of temperature, pain, body and limb movement, and body and limb position, are all mediated by the somatosensory system. Furthermore, touch can be subdivided into active touch (haptics) or passive touch. Haptics is critically dependent on propioceptive (from body and limb position) and kinaesthetic (from body and limb movement) feedback (Weisenberger, 2005). However, the experiments presented in this thesis utilise passive touch stimuli (where stimuli are presented passively to a stationary observer). For this reason, a detailed description of thermoreception (temperature), nociception (pain), propioception, kinaesthesia and haptics is not presented here; the interested reader may find information on these topics elsewhere (Craig & Rollman, 1999; Kruger, 1996).

In contrast to vision and audition, tactile stimuli can usually only be perceived when a stimulus is in direct contact with an observer. This is due to the nature of transduction in touch, which occurs within the skin in response to mechanical stimuli on the body surface. Humans possess three types of skin: hairy, glabrous and mucocutaneous; glabrous skin is only found on the palms of the hand and soles of the feet, whilst mucocutaneous skin is located at the junction with mucous membranes at the bodily orifices (Weisenberger, 2005). All types of skin are composed of two main layers. The epidermis is the superficial layer, and what we usually think of as ‘skin’ (Figure 1.9). Its keratinised surface forms a barrier against infection and injury. Beneath the epidermis lies the dermis, which is composed of connective tissue and capillary networks which supply the basement membrane of the epidermis. Sweat glands, sebaceous glands and hair follicles are all found within the dermis (Sherrick & Cholewiak, 1986). Beneath the dermis, subcutaneous tissue is composed mainly of adipose and connective tissues.

Tactile transduction is performed by *mechanoreceptors*, analogous to the retinal photoreceptors or inner hair cells of the cochlea. Four different types of mechanoreceptor cells have been putatively identified in humans: Meissner corpuscles, Pacinian corpuscles, Merkel discs, and Ruffini endings (Wolfe et al., 2006; Figure 1.9; Table 1.1).

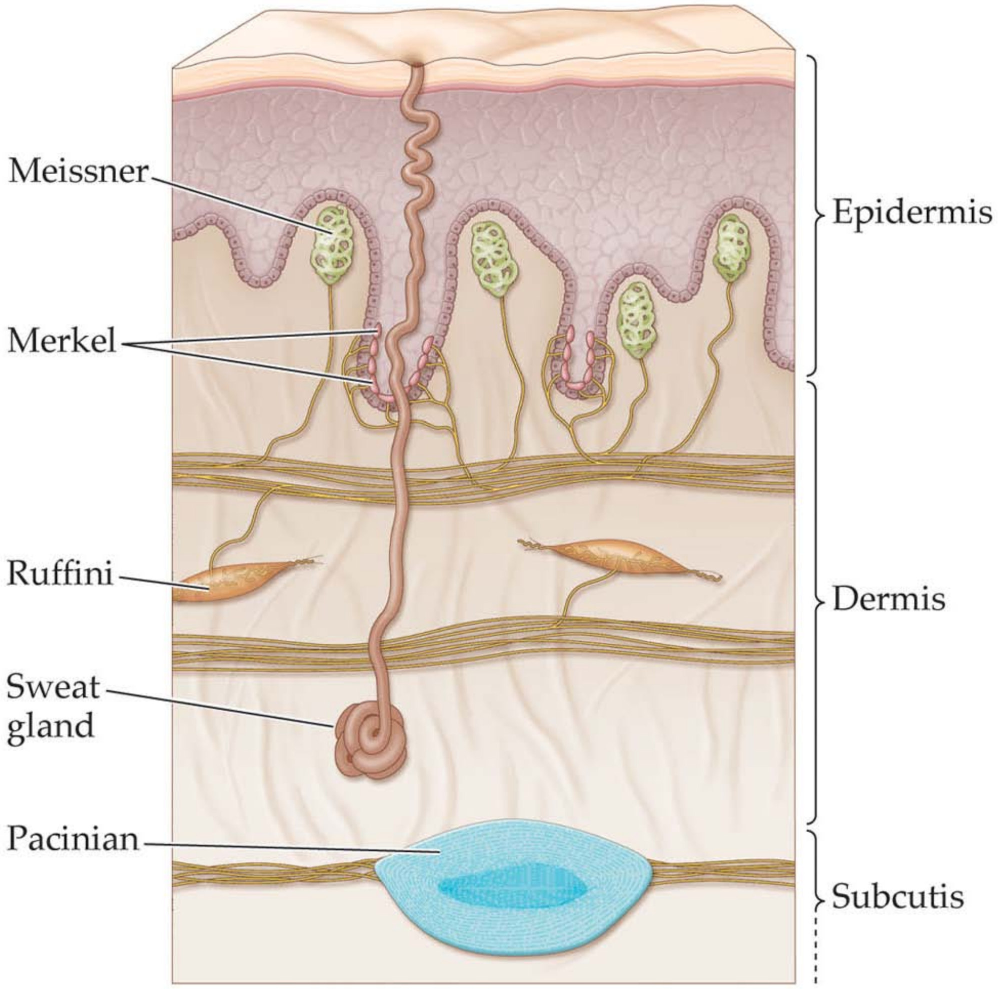


**Figure 1.9.** Diagram showing the four types of human mechanoreceptor and their location within the skin. Meissner corpuscles are located within the dermis, at the dermal papillae; Merkel discs at the epidermal basal cells; Ruffini endings in the upper layers of the dermis; and Pacinian corpuscles within the deep layers of the dermis (Wolfe et al., 2006).

The afferent nerve fibres associated with each type of mechanoreceptor can be classed according to the temporal dynamics of their response to mechanical stimulation and the size of their receptive fields (the area of skin upon which a tactile event will elicit a response from a particular afferent). In terms of response characteristics, tactile afferent fibres are classed as either *fast adapting* or *slow adapting* (Wolfe et al., 2006). Fast adapting fibres respond briskly at stimulus onset and offset but do not respond in between these events; Meissner and Pacinian corpuscles are both associated with fibres of this type. In contrast, slow adapting nerve fibres (associated with Merkel discs and Ruffini endings) respond continuously throughout the period of stimulation. In terms of receptive field size, tactile fibres terminating in Meissner corpuscles and Merkel discs are known to have small receptive fields in comparison with neurons associated with Pacinian corpuscles and Ruffini endings. These clear distinctions have led to a naming convention where tactile nerve fibres are classed as fast-adapting type I and II (FA I and FA II) and slow-adapting type I and II (SA I and II), with type I fibres having small receptive fields and type II large. For example, according to this classification tactile sensory nerve fibres terminating in Merkel discs are known as SA I fibres.

The four types of afferent nerve fibre may also be named according to their associated mechanoreceptors, e.g. an FA II fibre may also be referred to as a Pacinian fibre. To further complicate matters, some authors name the FA I system simply RA (from Rapidly Adapting), and the FA II system as PC (from Pacinian Corpuscles), whilst maintaining the SA I and SA II nomenclature (Hsiao, Johnson, & Yoshioka, 2003; Johnson, 2001; Johnson, Yoshioka, & Vega-Bermudez, 2000). For simplicity, in this brief introduction to tactile anatomy the FA I and II and SA I and II convention will be used throughout.

Like the different types of photoreceptors (rods and cones) in the eye, the different mechanoreceptors respond optimally to different stimulus characteristics. The spatial resolution, temporal resolution, and proposed functions of the four human mechanoreceptor types are summarised in Table 1.1. A fundamental difference between the tactile system in humans and monkeys is that monkeys are considered not to exhibit a SA II-type response (Hsiao et al., 2003; Weisenberger, 2005). This contrasts with the visual and auditory systems, which are generally considered to be extremely similar in humans and non-human primates.

| **Mechanoreceptor** | **Afferent neuron** | **Spatial resolution** | **Temporal sensitivity (Hz)** | **Proposed function/information signalled** |
|---|---|---|---|---|
| Meissner corpuscles | FA I | 20mm | 2-100 | Perception of tactile motion, adjustment of grip |
| Pacinian corpuscles | FA II | 3-5mm | 10-1000 | Detection of vibration |
| Merkel discs | SA I | 0.5mm | 0-100 | Perception of texture and form |
| Ruffini endings | SA II | 10mm | 0-20 | Perception of tactile motion/force direction, hand and finger position |

**Table 1.1.** Response characteristics and proposed functions of human mechanoreceptors. Adapted from Hsiao et al. (2003) and Johnson (2001).

Although much of the process of sensory transduction by mechanoreceptors is still poorly understood, it is believed that mechanical force applied to the mechanoreceptor surface opens ionic channels on the plasma membrane and initiates transmission of the nerve impulse (Fain, 2003). Neural signals from all four types of mechanoreceptor take the same route to the brain, a path shared with proprioceptive nerve fibres (nociceptive and thermoreceptive signals share a separate route to the brain). This route is called the dorsal-column-medial-lemniscal (DCML) pathway, or lemniscal pathway, and is illustrated in Figure 1.10. The only tactile afferent fibres not to travel the DCML pathway are those located in the head and face; these fibres project to sensory cortex through the cranial nerves (Weisenberger, 2005). The DCML pathway differs from the visual and auditory pathways in that rather than the two optic nerves and two auditory nerves, a total of 231 nerve trunks (posterior nerve roots) are distributed throughout the body (Regan, 1989), and because tactile neurons synapse initially in the spinal cord (Wolfe et al., 2006).

Tactile nerve fibres within the spinal column are segregated into two streams, with fibres from the lower body travelling in the gracile tract (fasiculus gracilis) and those from the upper body in the cuneate tract (fasciculus cuneatus) (Patestas & Gartner, 2006; Figure 1.10). The fibres synapse at the gracile nuclei (nucleus gracilis) and cuneate nuclei (nucleus cuneatus), respectively, of the medulla oblongata in the brainstem. At the medulla, the nerve fibres emerge decussated, i.e. fibres from the left side of the body travel to cortical areas in the right hemisphere, and ascend in the medial lemniscus pathway

(Patestas & Gartner, 2006). The cuneate and gracile nuclei have also been shown to project to the superior colliculus (Edwards et al., 1979). The next synapse in the ascending pathway is at the ventral posterior nuclei of the thalamus, and from here the nerve fibres project directly to somatosensory cortex in the brain via the corona radiata and internal capsule (Patestas & Gartner, 2006).

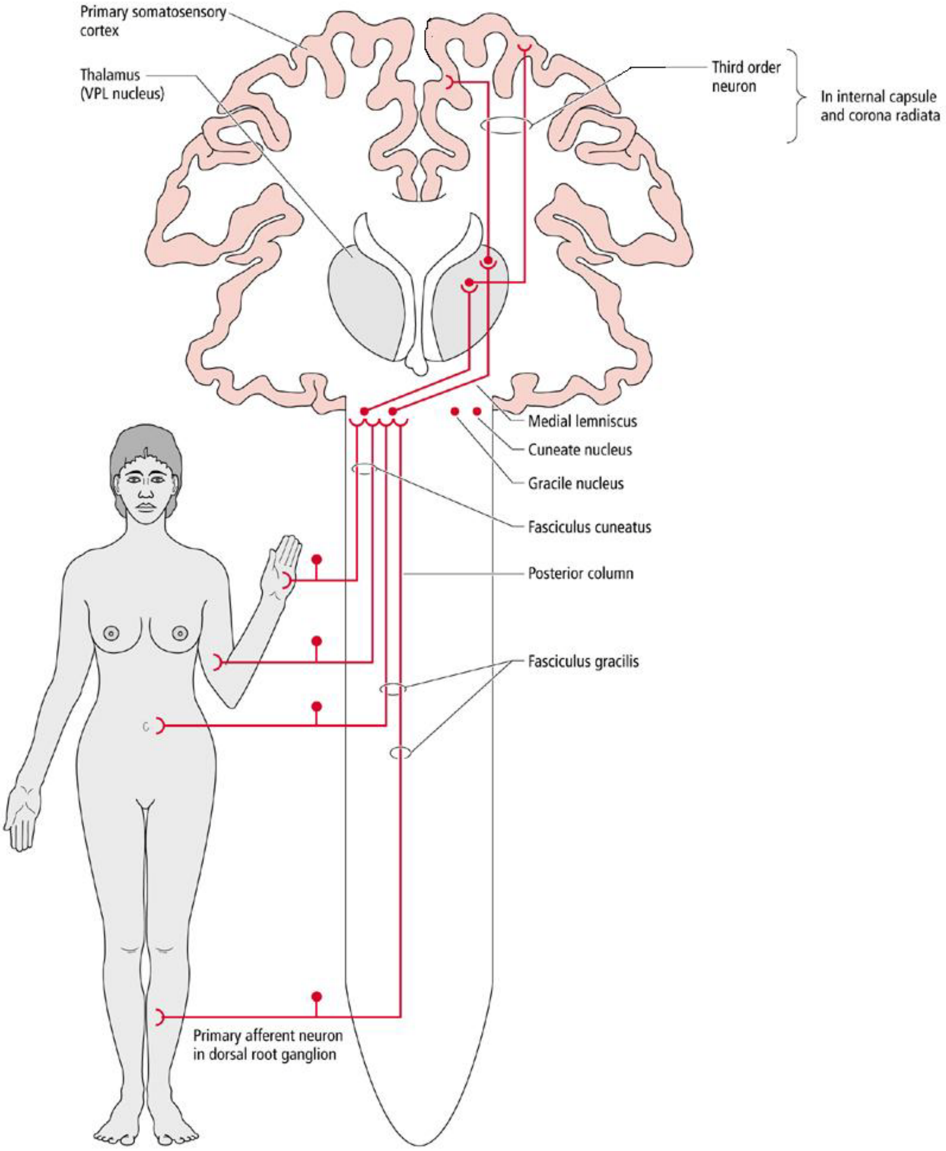


**Figure 1.10.** Diagram of the human ascending DCML pathway. Although the fibres of the fasciculus gracilis and fasciculus cuneatus are segregated, together they comprise the posterior column of the spinal cord (Patestas & Gartner, 2006).

Primary somatosensory cortex (S1; Brodmann areas 3a, 3b, 1 and 2) is located in the postcentral gyrus of the brain's parietal lobe. The majority of inputs project to area 3b (Sherrick & Cholewiak, 1986). Processing of tactile events is performed here and at related areas such as S2 and Brodmann areas 5 and 7. The location of somatosensory cortex is shown in Figure 1.11.

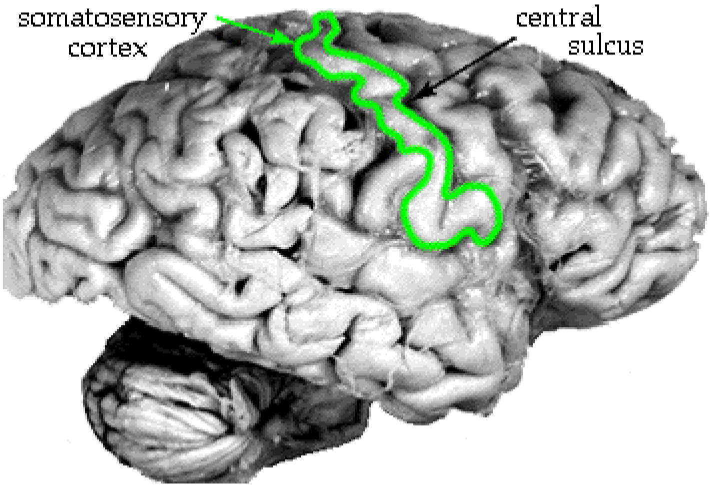


**Figure 1.11.** Diagram of the brain showing the location of somatosensory cortex. Taken from http://thalamus.wustl.edu/course/bassens.html.

As with visual and auditory events, the time taken for a tactile event to become available to the perceptual system can be divided into that taken for energy from an event to arrive at an observer, the delay caused by transduction of the stimulus at the receptor surface, and the time taken for the neural signal to reach sensory cortex. The first of these factors can be effectively discounted, as to stimulate the mechanoreceptors within the skin a tactile event must be in contact with the observer.

Mechanoreceptors located on the body surface of insects have been shown to produce a neural response approximately 100µsec following mechanical stimulation (Thurm, 1983), and intestinal Pacinian receptors in cat have been found to respond with a latency of 0.5-3.0msec (Gray & Sato, 1953). In humans, mechanoreceptor response time has been found to be 2.2msec on average (range 0.6-6.2msec) following air puff stimulation of the skin surface, with no significant differences in response latency being observed between the four types of mechanoreceptors (Mizobuchi et al., 2000; Mizobuchi et al., 2002). Note that this value incorporates time taken for skin deformation to be transmitted to the mechanoreceptors; the response latency of human mechanoreceptors in isolation is estimated at 0.5-0.7msec (Buchthal, 1982). This evidence suggests that mechanoreceptor transduction is rapid in comparison to the retinal transduction process, whilst being slightly slower than that at the cochlea.

Due to the spatial distribution of mechanoreceptors all over the body, the timing of cortical activity following tactile stimulation varies according to the location of the stimulus on the skin (e.g., stimulation of a toe vs. stimulation on the neck). Cortical latency also varies according to the size of the observer. The velocity of nerve impulses in tactile afferent nerve fibres located in the limbs has been measured at 44-80m/sec (Caruso et al., 1994; Macefield, Gandevia, & Burke, 1989; Mizobuchi et al., 2000), with no significant differences between velocities measured in the upper and lower limbs (Macefield et al., 1989). Assuming a conduction velocity value of 60m/sec (close to the mid-point of the range reported above), a tactile signal having to travel 0.8m further to S1 (i.e., the difference between a 1.8m adult and 1.0m child or monkey) would be expected to arrive approximately 13msec later.

Despite the potential confounds of stimulation site and observer size discussed above, it is still possible to obtain values for the latency of cortical activity following tactile stimulation. It has been demonstrated that early cortical components of the somatosensory evoked potential (SEP) peak at 10-25msec in alert macaque following direct electrical stimulation of the median nerve at the wrist (Arezzo, Vaughan, & Legatt, 1981; Peterson, Schroeder, & Arezzo, 1995). In humans, similar early SEPs peak between 20 and 50msec post-stimulus (Allison, Goff, Williamson, & van Gilder, 1980; Goldring, Aras, & Weber, 1970; Kelly, Goldring, & O'Leary, 1965; Stohr & Goldring, 1969). Human latency in S1 and S2 measured using MEG is approximately 20-30msec (Karhu & Tesche, 1999), which corresponds well with SEP latency values (Allison et al., 1980; Goldring et al., 1970; Kelly et al., 1965; Stohr & Goldring, 1969). All the values above are likely to underestimate cortical latency by approximately 2msec, as direct electrical stimulation of the median nerve was used as a proxy for conventional tactile stimulation. Such a method does not take account of the mechanoreceptor response latency and time required for mechanical vibration to travel through the skin. In summary, the latency of cortical activation following sensory stimulation of the somatosensory system appears to be intermediate to the values described for the auditory and visual systems. A summary of latencies in the visual, auditory and tactile systems is presented in Table 1.2.

| | Vision | Audition | Touch |
|---|---|---|---|
| **Receptor latency** | 7msec | 40µsec | 2.2msec |
| **Earliest cortical activation in humans**<br>* inferred from macaque data | 50-57msec* | 15-16msec* | 20-30msec |

**Table 1.2.** Summary of receptor discharge latencies and earliest activation of primary sensory cortices in humans following visual, auditory or tactile stimulation. References may be found in Sections 1.1.1, 1.1.2 and 1.1.3.

## 1.2 Introduction to multisensory processing

Although the preceding section describes the functional anatomy of the visual, auditory and tactile systems, it is obvious that many external events will be detected by more than one of these sensory systems. Where and how these differing sensory representations of the same external event interact within the nervous system is of clear interest. At the most fundamental level, this interaction is mediated by multisensory neurons.

A crossmodal or multisensory neuron may simply be defined as one which exhibits a response to events encompassing more than one sensory modality. They are found in all mammals at various stages in the neural processing hierarchy (Meredith, 2002; Stein & Meredith, 1993; Stein & Stanford, 2008). Higher regions of human cortex (beyond the level of primary sensory cortices) so far implicated as containing multisensory neurons are anterior superior temporal sulcus (STS), tempero-parietal association cortex (Tpt), ventral (VIP) and lateral (LIP) intraparietal areas of the parietal lobe, premotor and prefrontal cortex, and insular cortex (insula). These areas are thought to be ‘association areas’ where input from unimodal sensory cortices is combined to form a coherent multisensory percept of the environment (Calvert & Thesen, 2004). Proposed subcortical regions containing multisensory neurons include the superior colliculus (SC), inferior colliculus, claustrum, suprageniculate nuclei, medial pulvinar nuclei, and amygdala (Calvert & Thesen, 2004). A diagram showing these proposed multisensory areas within the human brain can be seen as Figure 1.12(a-c). It should be noted that multisensory neurons are unlikely to constitute the full range of neurons within these structures: for example, within SC (the most studied of the multisensory brain regions) multisensory neurons make up approximately 27% (in primates), 55% (in cat), and less than 10% (in hamster) of the total neuronal population (Meredith, 2002).

It is through multisensory neurons that the process of *multisensory integration* is facilitated. In the context of this brief overview, multisensory integration is described as a neurophysiological process; however, the term is also used in a psychophysical context to quantify behavioural outcomes when observers combine cues from different sensory modalities, e.g. (Alais & Burr, 2004; Andersen, Tiippana, & Sams, 2005; Ernst, 2007; Ernst & Banks, 2002; Ernst & Bulthoff, 2004; Roach, Heron, & McGraw, 2006). At the level of individual neurons, multisensory integration is defined as a statistically significant difference between the response evoked by a multisensory stimulus and the response evoked by the most effective unimodal component of this stimulus when presented individually (Meredith & Stein, 1983). This difference in response (relative to unimodal stimulus presentation) can take the form of *response enhancement* or *response depression*; response enhancement is characterised by an increase in neuronal firing, with response depression describing a reduction in firing rate. Cells exhibiting response enhancement appear to be more common than those exhibiting response depression, although some neurons may exhibit both enhancement *and* depression depending on the spatiotemporal relationship between stimuli (King & Palmer, 1985; Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1986).

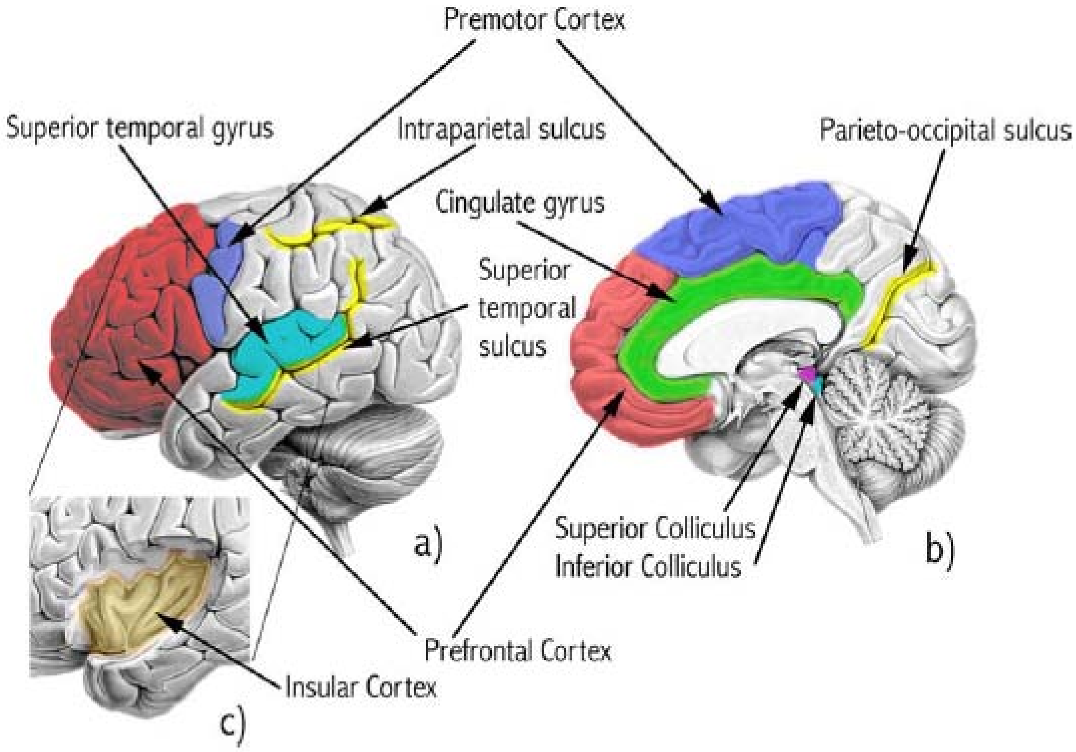


**Figure 1.12(a-c).** Diagram showing the location of proposed multisensory areas of the human brain, from (a) lateral and (b) mid-sagittal viewpoints. Figure 1.12 (c) shows the location of the insular cortex following temporal lobe dissection (Calvert & Thesen, 2004).

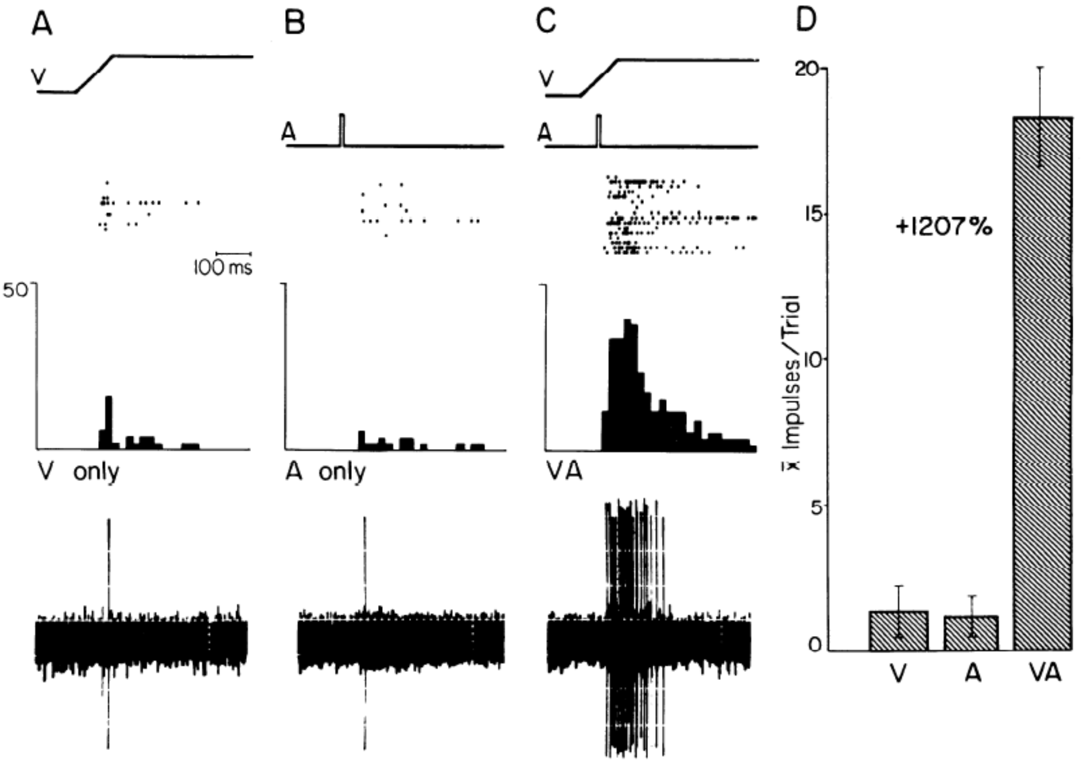


**Figure 1.13(a-d).** Illustration of superadditive/multiplicative multisensory integration in cat superior colliculus. Presented individually, visual and auditory stimuli (1.13 (a) and (b), respectively) produce minimal neural response, as represented by raster plots (second row), histograms (third row), and oscillograms (bottom row). The same stimuli presented in spatiotemporal correspondence (c) produce a significantly increased neural response which is quantified in (d) as a response enhancement of 1207% (Meredith & Stein, 1986).

Response enhancement may be further categorised as being *additive* or *subadditive*. Additive responses are characterised by being equal to the arithmetic sum of responses to the same stimuli presented unimodally, whilst subadditive responses are less than the sum of unimodal responses (Stein & Stanford, 2008). Frequently, the response is significantly greater than the sum of the unimodal responses; responses of this nature are termed *superadditive* or *multiplicative* (Meredith & Stein, 1986). An illustration of superadditive response enhancement is shown as Figure 1.13(a-d). Frequently, the degree of response enhancement is maximal when the responses to the unimodal stimuli are weakest (e.g., when the stimuli are of very low intensity). This phenomenon is known as *inverse effectiveness* (Meredith & Stein, 1986). An obvious ecological benefit of this is that two or more weak stimuli in spatiotemporal coincidence are more likely to be detected and acted upon appropriately.

As well as these multisensory areas, cortical areas traditionally considered unisensory (e.g., primary sensory cortices) have been shown to respond to heteromodal stimulation. This has been demonstrated not only in sensory-impaired (e.g., blind) subjects (Finney, Fine, & Dobkins, 2001; Ptito et al., 2008; Rao, Nobre, Alexander, & Cowey, 2007; Weeks et al., 2000), but also in normal human and animal populations. For example, recent animal studies have shown the existence of direct anatomical connections between visual and auditory (Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003), auditory and somatosensory (Cappe & Barone, 2005), and visual and somatosensory (Cappe & Barone, 2005; Fu et al., 2003) cortical areas. An implication of this is that multisensory interactions may be anatomically feasible at relatively early stages of neural processing, before processing at the high-level, association areas discussed above.

In normal humans, neuroimaging studies have been invaluable in illustrating the extent of these early-stage interactions between sensory modalities. Calvert and colleagues showed (using functional Magnetic Resonance

Imaging (fMRI)) that the auditory cortex is activated when observers viewed silent footage of speech (Calvert et al., 1997). A control condition using similar facial movements devoid of semantic context failed to show similar activation (Calvert et al., 1997). Subsequent work examined the different effects of visual, auditory and audiovisual (AV) speech on activation of visual and auditory cortex, and showed that the magnitude of activation induced by AV speech significantly exceeded the sum of unimodal activations (Calvert et al., 1999). This suggests that superadditive response enhancement is evident at the cortical level in addition to the subcortical.

Similar results have been found in the audiotactile (AT) and visuotactile (VT) domains. Simultaneous presentation of spatially aligned VT stimuli produced superadditive activation of the lingual gyrus, an area of visual cortex traditionally thought of as unisensory, as measured by fMRI (Macaluso, Frith, & Driver, 2000). Such an activation pattern has also been demonstrated in the parietal operculum, which contains secondary somatosensory cortical areas (Macaluso, Frith, & Driver, 2002). AT interactions at the cortical level have been demonstrated using MEG (Gobbelé et al., 2003; Lutkenhoner, Lammertmann, Simoes, & Hari, 2002), and it appears that at least some of these interactions may be subadditive in nature (Gobbelé et al., 2003).

Finally, fMRI data in humans has shown that activity in unimodal cortex can also be depressed by the presentation of stimuli in other modalities (Figure 1.14(a&b)). Presentation of visual stimuli depressed activity in auditory cortex (relative to a resting baseline), and vice versa (Laurienti et al., 2002). This activity was only recorded under conditions of unimodal stimulus presentation; when bimodal AV stimuli were presented, superadditive activation was observed in both visual and auditory cortices (Laurienti et al., 2002). Taken together, the results of these studies suggest that the range of neural mechanisms subserving multisensory integration in animal superior colliculus (sub- and superadditive response enhancement, response depression) can also be found within primary sensory cortex (traditionally thought of as unisensory) in normal humans.

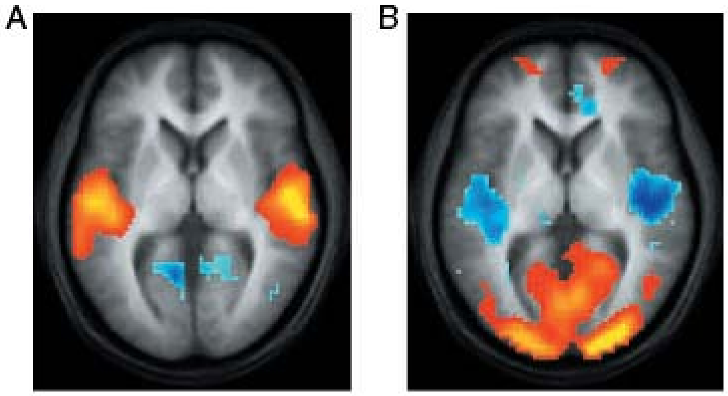


**Figure 1.14(a&b).** Activation (red/orange) and deactivation (blue) of cortical areas caused by auditory (a) and visual (b) stimulation. Following an auditory event, auditory cortex is activated and extrastriate visual areas (lingual, fusiform and posterior cingulated gyrii) are deactivated. Following a visual event, visual cortex is activated and auditory cortex (superior, middle and inferior temporal gyrii) is deactivated (Laurienti et al., 2002).

### 1.2.1 Latency of multisensory brain areas

The majority of multisensory neurophysiological studies have investigated processing within the superior colliculus (SC). As well as being part of the subcortical visual pathway (Section 1.1.1), the SC receives both ascending and descending input from a large number of visual, auditory and somatosensory structures (Edwards et al., 1979; Huerta & Harting, 1984). In cats, descending connections from the anterior ectosylvian sulcus (AES) and lateral suprasylvian cortex (rLS) are thought to be particularly important in integrating information from different modalities; when these brain areas are deactivated, multisensory integration in SC neurons is inhibited or abolished (Alvarado, Stanford, Vaughan, & Stein, 2007; Jiang, Jiang, & Stein, 2002; Jiang, Wallace, Jiang, Vaughan, & Stein, 2001).

The ascending *and* descending nature of inputs to the SC implies that multisensory neurons in this structure may be activated at a range of latencies depending on the neural origin of the incoming sensory information. It is known that multisensory neurons in cat SC respond on average 83, 19 and 27msec following visual, auditory and tactile stimulation, respectively (Meredith et al., 1987). Visual and auditory data obtained from guinea pigs are comparable to these values (King & Palmer, 1985). These latencies are slightly longer than those recorded in the primary sensory cortices of monkeys (Sections 1.1.1-1.1.3), despite the physically smaller size of cats and guinea pigs. Although it is known that subcortical sensory input to SC (e.g., from the subcortical visual pathway) predominantly projects to the multisensory neuron-containing deeper laminae (Edwards et al., 1979), the timescale of activation may reflect the importance of descending AES and rLS sensory input in multisensory integration. In other words, it is possible that the necessity of *descending* sensory input from AES and/or rLS (or their primate homologues) may delay the response of SC multisensory neurons to *ascending* sensory signals.

Also in cats, multisensory SC neurons have been reported to discharge on average 99msec following auditory stimulation (Rowland, Quessy, Stanford, & Stein, 2007). However, AV stimulation resulted in average discharge latencies of 80msec (Rowland et al., 2007), which represents a statistically significant speeding of multisensory-evoked activation relative to fastest unimodal-evoked. The reason for the difference in unimodal auditory latency in this study with that reported previously (Meredith et al., 1987), 99 vs. 19msec, is unclear; however, this does not affect Rowland et al.'s conclusion that multisensory integration in SC shortens response latencies as the same animals and individual neurons were used to record both unimodal and AV data sets. A similar pattern was observed when investigating AV, AT, VT and trimodal integration in cat basal ganglia (Nagy, Eordegh, Paroczy, Markus, & Benedek, 2006). Smaller (but still significant) reductions in response latency following AV stimulation relative to unimodal visual stimulation have also been noted in V1 neurons in monkey (Wang, Celebrini, Trotter, & Barone, 2008), although the lack of any unimodal auditory condition for comparison represents a weakness of this study.

Of the multisensory brain regions other than SC, the Superior Temporal Sulcus (STS) complex is relatively well-studied because of its putatively important role in AV speech processing. As measured with electroencephalography (EEG), latency of this area following congruent (matching) AV speech in humans is approximately 65-70msec, with incongruent speech resulting in a slight (~10msec) delay in activation

(Bernstein, Auer, Wagner, & Ponton, 2008). At the behavioural level (as measured by reaction time (RT)), this latency advantage for congruent speech relative to incongruent is also apparent (Murase et al., 2008). Unimodal somatosensory-evoked activation in the same area (following median nerve stimulation at the wrist) occurs on a similar timescale, potentially allowing multisensory integration in this area relatively early post-stimulus (Tesche, 2000).

Unfortunately, studies examining latency in the remaining multisensory brain areas appear to be uncommon. In intraparietal sulcus (part of the parietal lobes), activity following AV speech has been demonstrated at latencies as short as 55-65msec post-stimulus; as in STS, congruent speech results in slightly reduced latency relative to incongruent (Bernstein et al., 2008). Although the existence of multisensory neurons in human tempero-parietal association cortex (Tpt) has been confirmed via intracranial recordings (on patients undergoing surgery for intractable epilepsy), the latencies were found to vary widely between observers (Matsuhashi et al., 2004). Unfortunately, neural recording sites and the range of experimental conditions undertaken varied widely between observers (n = 6) in this study, making meaningful interpretation of latency data problematic. However, the authors suggest that the generally late timescale of activation is strongly consistent with the proposed secondary/integratory function of Tpt (Matsuhashi et al., 2004).

Finally, in macaque VIP, visual and tactile events cause neuronal firing an average of 85 and 42msec post-stimulus, respectively (Avillac, Ben Hamed, &

Duhamel, 2007). Employing the temporal scaling factor discussed in Section 1.1.1 (Schroeder et al., 2004; Schroeder et al., 1995) suggests human VIP latencies of approximately 142 and 70msec. In contrast to results in cat SC (Rowland et al., 2007) and basal ganglia (Nagy et al., 2006), multisensory integration of VT signals in macaque VIP appears *not* to reduce latency beyond that of the fastest unimodal signal. Instead, multisensory latency appears to lie intermediate to visual and tactile unimodal values at 69msec (Avillac et al., 2007), suggesting that similar integration occurs approximately 115msec post-stimulus in humans. Although VT integration in macaque VIP appears to offer no latency benefits, multisensory neurons here display both sub- and superadditive response enhancement as well as response depression (Avillac et al., 2007). When considered along with data in animal SC (Meredith et al., 1987), basal ganglia (Nagy et al., 2006), and human primary sensory cortices (Calvert et al., 1999; Gobbelé et al., 2003; Macaluso et al., 2000), this suggests that multisensory integration displays very similar neurophysiological characteristics throughout the central nervous system.

### 1.2.2 Temporal factors in multisensory integration

Multisensory integration at the level of single neurons is critically dependent on the spatial (Meredith & Stein, 1986) and temporal (Meredith et al., 1987) configuration of stimuli. Meredith et al. (1987) investigated the temporal tolerance of multisensory integration in cat superior colliculus using spatially-aligned auditory, visual and tactile stimuli. Presentation of bimodal stimuli significantly increased the number of nerve impulses discharged, the duration of the discharge period, and rate of discharge relative to unimodal presentations. The degree and pattern of multisensory integration for individual neurons varied with the temporal relationship between stimuli: the majority (51%) of tested neurons displayed varying degrees of response enhancement over the range of temporal disparities investigated, 22% displayed varying degrees of response depression, and 27% displayed both enhancement *and* depression at different temporal disparities (Figure 1.15).

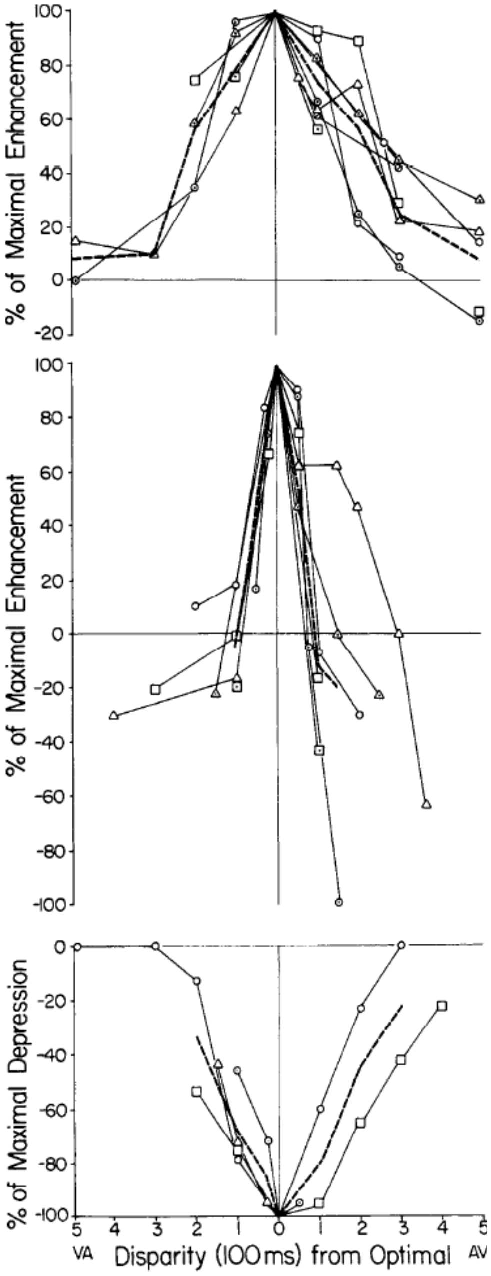


**Figure 1.15.** Response profiles of multisensory neurons in cat superior colliculus following presentation with spatially congruent audiovisual events at varying temporal disparities. Symbols connected by solid lines represent the responses of a single neuron over a range of temporal disparities (Stimulus Onset Asynchronies (SOAs)). The top panel shows a population of neurons that exhibit response enhancement when presented with simultaneous audiovisual events, with responses declining with increasing temporal disparity between the stimuli. The bottom panel shows a population of cells which exhibit response depression when presented with identical stimuli, with the depression reducing with increasing asynchrony between the stimuli. The middle panel shows a population of neurons which exhibit both response enhancement *and* response depression, dependent on the temporal relationship between the two stimuli. Note that for all neurons illustrated, the magnitude of the enhancement or depression is maximal when the stimuli are presented physically simultaneously (Meredith et al., 1987).

In addition to these results, it was found that multisensory integration tended to be maximal when bimodal stimuli were presented approximately simultaneously (Meredith et al., 1987). Given the differences in transduction and neural latencies discussed previously (Sections 1.1.1-1.1.3), this may be considered surprising – for example, the neural representation of a proximal auditory event would be expected to reach sensory cortex 40-50msec before that of a physically simultaneous auditory event. Within cat superior colliculus, this disparity appears to be even more pronounced; as previously discussed, the mean latency of SC multisensory neurons to unimodal visual events is 83msec, compared to 19msec for auditory events and 27msec for tactile events (Meredith et al., 1987).

The most likely way for the nervous system to successfully integrate signals from different modalities (and hence with different latencies) is via the extended discharge trains characteristic of multisensory neurons. Essentially, multisensory integration is likely to be physiologically mediated by temporal overlap of unimodal discharge trains, rather than simultaneous arrival of sensory signals (with different latencies) at multisensory neurons. It is suggested that such a mechanism would permit multisensory integration over a wider range of distances than would be possible were integration based purely on stimulus latency (Stein et al., 2005), and that multisensory integration is greatest when the peaks of the discharge trains temporally coincide (Stein & Stanford, 2008). This is illustrated in Figure 1.16.

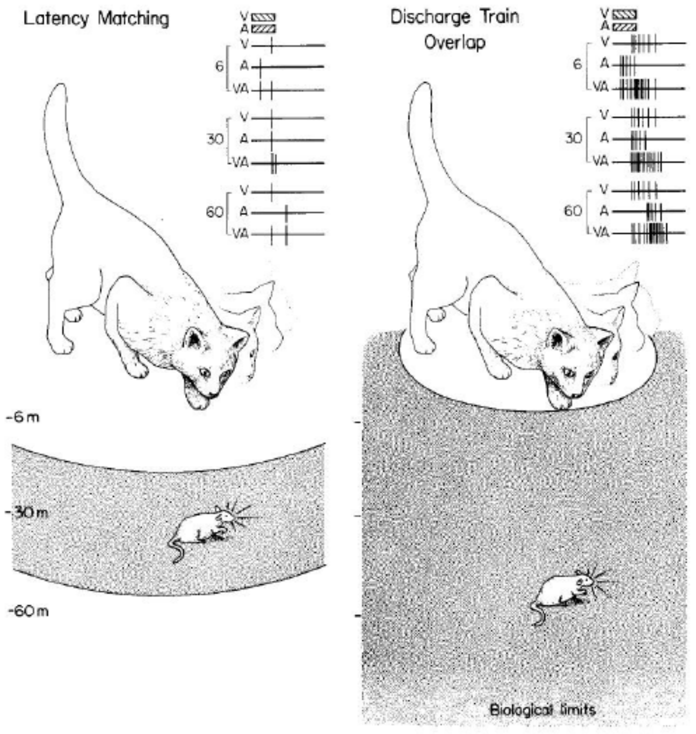


**Figure 1.16.** Ecological advantages of multisensory (audiovisual) integration mechanisms based on overlapping discharge trains over integration based on latency matching of unimodal signals. In the latter case (left diagram), the difference between visual and auditory latencies at multisensory neurons of cat superior colliculus mean that multisensory integration is possible only over a range of approximately 7-39m due to differences in neural latency and the velocities of light and sound in air. Conversely, integration based on overlap between unimodal discharge trains (right diagram) would allow integration over a much wider range of neural latencies, and hence distances. Note that integration is also dependent on spatial correspondence between unimodal stimuli (Meredith et al., 1987).

Although the integration mechanism described above and in Figure 1.16 has clear ecological advantages, it implies that *any* auditory and visual events occurring within 7-39m from the animal will be integrated within the nervous system. This would represent a disadvantage, as unrelated auditory and visual stimulation could be erroneously perceived as sharing a common cause. However, other work which has found that spatial correspondence between external events is necessary for multisensory integration (Meredith & Stein, 1986) renders this outcome less likely; for example, visual and auditory events sharing approximate spatiotemporal correspondence are more likely to relate to a common cause than those in temporal correspondence and spatial discordance.

Temporal correspondence between AV stimuli has also been shown to influence multisensory integration in human cortical areas. In STS as well as extrastriate visual and auditory areas, simultaneous presentation of AV speech sounds resulted in maximal multisensory integration as measured with fMRI (van Atteveldt, Formisano, Blomert, & Goebel, 2007). Additionally, using unpredictable streams of visual and auditory stimuli (also with fMRI), both response enhancement and response depression (relative to the activation caused by unimodal presentation of the stimuli) was demonstrated within STS depending on the temporal relationship between the streams (Noesselt et al., 2007). The use of simple, non-verbal stimuli in this study supports the notion of a wider role for AV integration in STS than processing of speech.

Within macaque VIP, multisensory VT integration has also been shown to depend on the temporal relationship between stimuli: as described above in cat SC with AV stimuli, integration is maximal when stimuli are presented simultaneously, despite significant differences in latencies of the unimodal stimuli (Avillac et al., 2007). The effect of VT temporal relationship on neuronal firing rate is shown in Figure 1.17.

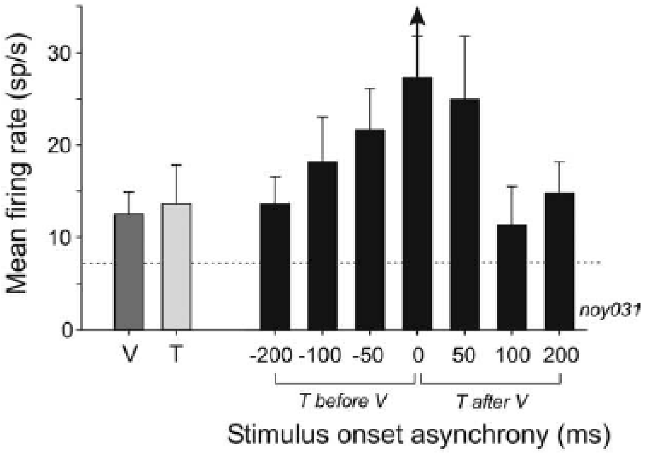


**Figure 1.17.** Effect of temporal relationship between visual and tactile stimuli on neural firing rate in monkey VIP. Firing rate elicited by the visual and tactile stimuli presented individually is indicated by the 'V' and 'T' bars on the left. Baseline firing rate is shown by the dashed horizontal line. When both stimuli are presented, firing rate is maximally increased when the stimuli are simultaneous (stimulus onset asynchrony = 0), with increasing temporal discordance reducing the observed firing rate (Avillac et al., 2007).

Although there appears to be a wide range of neurophysiological evidence for multisensory integration in animals published since the early work of Stein and Meredith in 1983, it is only recently that this work has started to be appraised more critically. Currently, particular attention seems to be dedicated to examining the principle of inverse effectiveness first described by Meredith and Stein (1986). As already discussed, this rule states that response enhancement of a multisensory stimulus is maximal when the neural responses to the unimodal stimuli are weakest (e.g., when the stimuli are of very low intensity). Recent work has suggested that the evidence supporting inverse effectiveness may in fact be attributable to the statistical phenomenon of regression toward the mean and the choice of statistical analysis performed on experimental data (Holmes, 2007, 2008); in other words, it is hypothesised that inverse effectiveness may represent a statistical artefact rather than a neurophysiological phenomenon. Further, it has been shown that results consistent with inverse effectiveness may be obtained when using *random* numbers to simulate unisensory and multisensory neural firing rates, which suggests that inverse effectiveness may be a product of the random activity or ‘noise’ (Section 3.1.1) present in all neurons (Holmes, 2008). Clearly, future studies demonstrating conclusive behavioural evidence (or lack of evidence) for inverse effectiveness will make a valuable contribution to this emerging debate.

# Chapter 2

## 2.1 Psychophysical measures of temporal processing

Thus far, I have examined the time course of the sensory transduction process in the visual, auditory and tactile modalities (Sections 1.1.1-1.1.3). Although undoubtedly important when considering the temporal processing characteristics of the sensory modalities, this is an anatomical and physiological, rather than perceptual, estimate of the time course of sensory processing.

In this context, it is important to consider what is meant by the term perceptual in psychological and philosophical research, and to differentiate perception from sensation. Typically, sensation refers to the ability of sensory systems to detect various forms of energy – for example, the ability of the visual system to detect light, the auditory system sound waves, and the tactile system to detect skin contact. However, sensation alone is meaningless without processing and interpretation by the central nervous system to convey meaning on the raw sensory information (Pike & Edgar, 2005). Through this processing and interpretation, the accumulated prior experience of an observer influences the perceptual outcome. One example of this may be found in a study which examined the influence of memory on colour perception; observers asked to adjust an image of a banana so that the image appeared grey found that when objective achromaticity (grey) was reached the banana was still perceived as yellow; in order to appear perceptually yellow in colour, the image was required to be objectively blue (the opponent colour of yellow) (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006). In this case, observers' knowledge of bananas typically being yellow caused them to perceive a grey image of a banana as being slightly yellow in colour.

It can therefore be seen that perception is not simply dependent on 'bottom-up' sensation, but also on 'top-down', cognitive influences (Laming, 1997). Whilst Sections 1.1.1 – 1.1.3 described the mechanisms of visual, auditory and tactile sensation and their time course, in this chapter I review purely perceptual aspects of temporal processing within and between the sensory modalities.

### 2.1.1 Reaction time

The most fundamental perceptual measure of sensory latency is the simple reaction time (RT). This may be defined as the time taken by an organism to detect and respond to a stimulus in the absence of any cognitive demand. RT is perhaps the oldest measure of perceptual latency used in psychology, and has been studied extensively since the mid-nineteenth century (Helmholtz, 1850).

Although substantial variations exist between studies and observers, normal RT values are usually considered to be approximately 160-250msec for vision (Arrighi, Alais, & Burr, 2005; Colavita, 1974; Diederich & Colonius, 2004;

Elliott, 1968; Gielen, Schmidt, & Vandenheuvel, 1983; Hershenson, 1962; Robinson, 1934; Rutschmann & Link, 1964; Stone et al., 2001; Todd, 1912), 120-200msec for audition (Arrighi et al., 2005; Colavita, 1974; Diederich & Colonius, 2004; Elliott, 1968; Gielen et al., 1983; Hershenson, 1962; Kohfeld, Santee, & Wallace, 1981; Robinson, 1934; Stone et al., 2001; Todd, 1912), and 155-220msec for touch (Diederich & Colonius, 2004; Harrar & Harris, 2005; Robinson, 1934; Todd, 1912). A potential confound peculiar to tactile RT is that the time taken for the sensory signal to reach the brain varies according to the distance of the stimulated bodily region from the brain (Section 1.1.3). As seen in Figure 2.1, this results in a linear increase in tactile RT with increasing distance between the bodily region stimulated and the brain (Harrar & Harris, 2005).

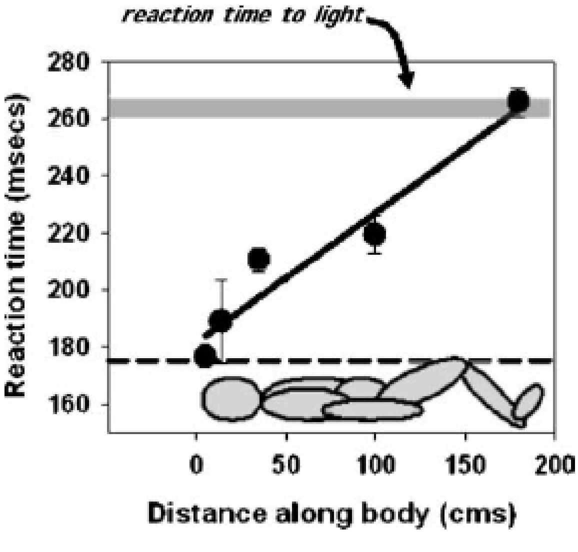


**Figure 2.1.** Diagram showing linear increase in tactile RT with increasing distance between bodily region stimulated and the head (Harrar & Harris, 2005).

The pattern of RT values in the visual, auditory and tactile domains corresponds well with the physiological and neural latencies discussed in Chapter 1, in that auditory RT appears to be fastest, followed by tactile, with visual being the slowest. Auditory RT is considered to be faster than visual by

40-50msec (Arrighi et al., 2005; Boulter, 1977; Elliott, 1968; Goldstone, 1968; Hershenson, 1962; Rutschmann & Link, 1964). Tactile RT is considered to lie between visual and auditory values (Boulter, 1977; Robinson, 1934; Todd, 1912), but when distal bodily regions (e.g., a toe) are stimulated tactile RT may be slower than visual (Diederich & Colonius, 2004). The different RT values in different sensory modalities are thought to reflect differences in the time course of peripheral transduction processes (Sections 1.1.1-1.1.3), rather than in the cortical processing of sensory events (Brebner & Welford, 1980; Elliott, 1968).

From the manner in which the relationship between RT values mirrors the relationship between physiological and neural latencies, we can infer that the time taken to execute a motor response to a stimulus is similar in the three modalities. This is consistent with the experimental finding that a common motor mechanism is used to respond to visual, auditory and tactile events in a simple RT task (Weeks, Honda, Catalan, & Hallett, 2001). Subsequent work has additionally suggested that a common neural network is activated during visual, auditory and tactile simple RT tasks: in addition to modality-specific activation of primary sensory cortices, activation in right posterior superior temporal cortex, right and left premotor cortex, right occipitotemporal gyrus, and medial frontal gyrus/supplementary motor area was noted (Kansaku, Hanakawa, Wu, & Hallett, 2004). It is proposed that these areas are involved in detection of sensory cues and executing a motor response; the same network is activated regardless of sensory input modality or motor output effecter (Kansaku et al., 2004). The existence of such shared networks is likely to explain the close correspondence between physiological/neural latency and simple RT.

Simple RT in the visual, auditory and tactile modalities is strongly affected by stimulus intensity (Figure 2.2), in that RT to a weak stimulus is slower than that to a strong stimulus (Cattell, 1886; Diederich & Colonius, 2004; Jaskowski & Sobieralska, 2004; Kammer, Lehr, & Kirschfeld, 1999; Kohfeld, 1971; Kohfeld et al., 1981; Pins & Bonnet, 2000; Raab, 1962a; Ulrich, Rinkenauer, & Miller, 1998). Whether the speeding of RT with increasing stimulus intensity reflects processes within the sensory pathways or within the brain has been debated extensively, as summarised by Nissen (1977). However, it is clear that increased stimulus intensity causes increased activity within primary sensory cortices (Klingaman & Anch, 1972; Tepas & Armington, 1962), which is usually considered to reflect processes within the sensory pathways (Nissen, 1977). Intensity effects on RT are therefore likely to reflect response enhancement at this peripheral level.

RT is also affected by stimulus duration (Figure 2.2), with longer durations leading to reduced RT up to a critical duration where RT asymptotes (Froeberg, 1907; Hildreth, 1973; Ulrich et al., 1998). This effect appears to be less pronounced than the intensity effect discussed above. Within the visual domain, RT is also affected by the size of the stimulus, with larger stimuli causing reduced RT (Ferree & Rand, 1927; Froeberg, 1907). Similar effects are observed when comparing binocular to monocular stimulus presentation – a clear speeding of RT is apparent under conditions of binocular presentation

(Blake, Martens, & Di Gianfilippo, 1980). Finally, visual RT has also been shown to vary according to retinal location stimulated (Rains, 1963) and stimulus chromaticity (McKeefry, Parry, & Murray, 2003; Nissen & Pokorny, 1977).

The processing of stimuli in a simple RT task has been mathematically modelled by Miller and Ulrich in a way that accounts for the effects of stimulus intensity, duration and area discussed above (Miller & Ulrich, 2003). This theory, entitled the *parallel grains model*, assumes that stimulus presentation activates a number of codes, or *grains*, within the perceptual system; the number of activated grains on any trial varies according to the intensity, duration and size of the stimulus. Each grain 'races' to a central decision centre, with the speed of each grain being random due to the omnipresent noise within the sensory system (Section 3.1.1). Detection of the stimulus is accomplished when a set number of grains (corresponding to observers' decision criterion) reach the decision centre. When this occurs, the decision centre signals to the motor system to respond to the stimulus (Miller & Ulrich, 2003). The execution of the motor response is assumed to take some time which remains constant even whilst RT varies; this assumption is consistent with older experimental evidence (Botwinick & Thompson, 1966).

In essence, the parallel grains model assumes that speeded RT in response to more intense (or longer) stimuli represents a form of statistical facilitation. For example, a longer (or more intense) stimulus is proposed to increase the probability of any individual grain being activated. The more grains activated (whilst criterion remains constant), the shorter the average time required for the number of grains corresponding to criterion to reach the decision centre (Miller & Ulrich, 2003). Statistical facilitation therefore increases with activation of increasing numbers of grains. Predictions of the parallel grains model are compared with experimental data in Figure 2.2.

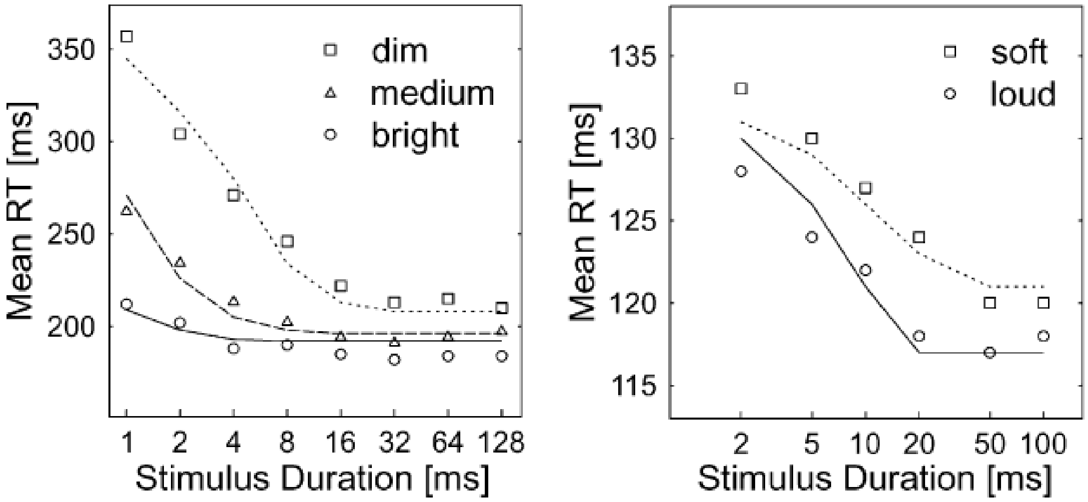


**Figure 2.2.** Comparison of RT results predicted with the parallel grains model (Miller & Ulrich, 2003) with experimental data. The points on the left graph show visual RT as a function of stimulus duration for three levels of stimulus intensity, and were obtained from Hildreth (1973). The lines represent the RT predictions of the parallel grains model. The points on the right graph show similarly-plotted auditory RT obtained from Raab (1962a), with the lines also showing the prediction of the parallel grains model (Miller & Ulrich, 2003).

It has also been shown that RT to stimuli in a given modality may be speeded by the concurrent presentation of stimuli in one or more other modalities, a phenomenon known as the *redundant signals effect* (RSE). These effects are measured when attention is divided between the modalities and observers are instructed to respond to signals in only one modality. Results typically show that unimodal RT > bimodal RT > trimodal RT (Diederich & Colonius, 2004;

Hecht, Reiner, & Karni, 2008; Todd, 1912). Miller has proposed that for the RSE to occur, signals from different modalities must be combined at some stage *prior* to the initiation of a motor response, and that the response is only initiated when the combined sensory signals exceed criterion (Miller, 1982). This assumption is easily incorporated into the parallel grains model by assuming that stimuli in each modality activate separate populations of grains (Miller & Ulrich, 2003). Thus, even though observers are required to respond only to stimuli presented in one modality, bi- or trimodal events activate a larger total number of grains. The more grains activated (whilst criterion remains constant), the shorter the average time required for the number of grains corresponding to criterion to initiate the motor response (Miller & Ulrich, 2003).

Alternatively, it has been proposed that in such studies, signals in the different sensory modalities are processed simultaneously and in parallel, with each modality effectively engaging in a ‘race’ to generate a motor signal (Raab, 1962b). The observer would therefore respond to whichever of the stimuli generated the motor signal first. Processing within each modality is thought to be independent of that in the other modality(s) (Meijers & Eijkman, 1977). Although race models (Meijers & Eijkman, 1977; Raab, 1962b) are more parsimonious than the parallel grains model described above, they have been shown to underestimate the magnitude of the RSE as measured experimentally (Miller, 1982). The parallel grains model is therefore more appealing as it accounts for the effects of stimulus intensity, duration and size in a simple (rather than bi- or trimodal) RT context, as well as the RSE, in a manner consistent with experimental data (Miller & Ulrich, 2003).

As well as simple RT, which requires no cognitive demand of the observer, a choice RT task may be employed. In this paradigm, observers must choose their responses according to the stimuli presented on a trial-by-trial basis. For example, observers may have to press different keys in order to respond to stimuli presented in different modalities (Colavita, 1974; Hartcher-O'Brien, Gallace, Krings, Koppen, & Spence, 2008; Koppen & Spence, 2007c) or at different locations (Spence, Nicholls, & Driver, 2001a). The additional cognitive demand of the task (modality discrimination followed by response selection) results in choice RT values being longer than simple RT to identical stimuli (Colavita, 1974; Dittrich & Henderson, 1999; Hohnsbein, Falkenstein, Hoormann, & Blanke, 1991). As opposed to simple RT, which seems to be dependent on the speed of the peripheral transduction mechanisms in the various modalities, choice RT is dependent on the task performed by observers and is therefore strongly influenced by high-level, cognitive factors. For example, Colavita (1974) reported that visual choice RT was approximately the same as auditory choice RT, despite the clear processing latency advantage of auditory stimuli (see preceding discussion and Sections 1.1.1 and 1.1.2).

### 2.1.2 Temporal resolution

A fundamental property of any sensory system is its temporal resolution, or the minimum time between stimuli such that the stimuli are perceived as being separate events. When two flashes of light (presented at the same location) are used as stimuli, this value ranges from 15-79msec (Exner, 1875; Kietzman, 1967; King, 1962; Lewis, 1967), dependent on stimulus intensity and duration (Kietzman, 1967; Lewis, 1967). This value represents the two-flash flicker (TFF) threshold (Herrick, 1974). However, within the visual domain it is more common to measure observers' ability to resolve a stream (rather than a pair) of stimuli as being separate. Such a task quantifies the Critical Flicker Fusion frequency (CFF) of an observer, or the maximum temporal frequency of a flickering light which can be reliably discriminated from a constant (non-flickering) light. A light flickering at a frequency higher than the CFF appears to be constant (Sekuler & Blake, 1994).

In humans, the maximum CFF is 40-60Hz, i.e. 40-60 on/off cycles per second (Davson, 1972; King, 1962; Landis, 1954; McFarland, Warren, & Karis, 1958), equivalent to a TFF of 17-25msec. Thus, it appears that observers are more sensitive to streams (rather than pairs) of flickering stimuli when measured using the same experimental apparatus (King, 1962). This has been ascribed to simple probability summation in that observers are more likely to perceive two stimuli as temporally distinct given repeated presentations, as in a CFF experiment (Herrick, 1974). It has also been proposed to reflect a decrease in visual persistence with repetitive streams of identical stimuli (Pariyadath & Eagleman, 2008).

CFF varies according to stimulus intensity (Ferry, 1892; Porter, 1902), stimulus contrast (Kelly, 1961), size of the retinal area stimulated (Granit & Harper, 1930), and the age of the observer (McFarland et al., 1958). The relationship between luminance and CFF is quantified by the Ferry-Porter law, which states that a logarithmic increase in stimulus luminance results in a linear rise in the CFF (Ferry, 1892; Porter, 1902), whereas the Granit-Harper law states that a logarithmic increase in retinal area stimulated causes a linear rise in CFF (Granit & Harper, 1930).

A consequence of the temporal properties of the visual system quantified by the Ferry-Porter and Granit-Harper laws is that CFF measured with a target of constant size and luminance is lower at the retinal periphery than at fixation (Hartmann, Lachenmayr, & Brettel, 1979; Hecht & Verrip, 1933). However, subsequent work has shown that when stimulus size and luminance is scaled so that the number of stimulated retinal ganglion cells and incident luminous flux is constant across the retina, CFF does not vary according to retinal location (Rovamo & Raninen, 1984). This finding led to revision of the Granit-Harper and Ferry-Porter laws to reflect that the number of ganglion cells stimulated and the luminous flux they collect determines CFF, rather than simply illuminance and retinal location in isolation (Rovamo & Raninen, 1988).

In audition, the minimum temporal gap between two stimuli with which the stimuli are perceived as separate is 5-10µsec (Leshowitz, 1971). In this study, single and paired stimuli were used; when analysed, these stimuli were found to differ in their spectral content, with the paired stimulus containing high frequency content reduced in intensity (Figure 2.3). These spectral differences acted as an additional cue to aid discrimination of the two stimuli (i.e., the task resembled one of spectral discrimination in addition to temporal discrimination), and when high frequencies were masked by the use of a low-pass filter resolution thresholds rose significantly (Leshowitz, 1971). Other studies using stimuli without this spectral cue obtained resolution threshold values ranging from 1-4msec (Corso, 1980; Exner, 1875; Gescheider, 1966, 1967; Green, 1971).

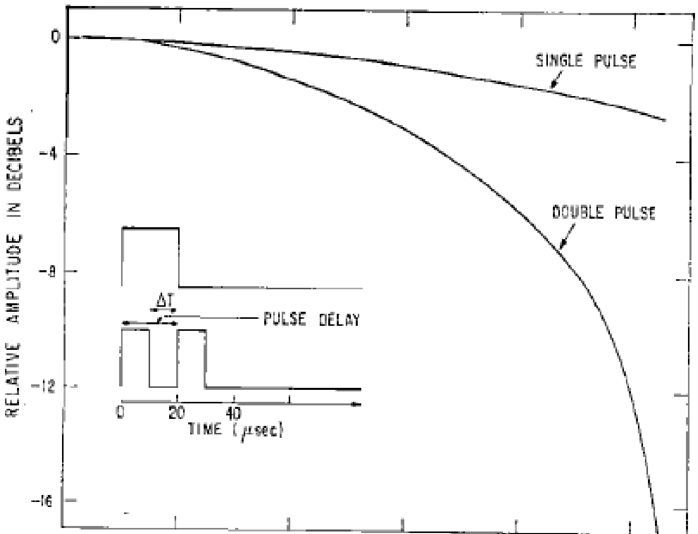


**Figure 2.3.** Spectral content of the single and double auditory stimuli used by Leshowitz (1971). The single stimulus contains more high frequency content than the double stimulus (which consists of two clicks separated by an interval Δt); with the aid of this cue, observers could distinguish between the single and paired stimuli when Δt was 5-10µsec (Leshowitz, 1971).

Auditory temporal resolution may also be measured using a gap-detection task; this is typically done using a temporal two-alternative forced choice paradigm (2AFC; Section 3.1.2), where observers are forced to choose which one of two bursts of noise contains a brief gap. The smallest size of the gap which can reliably discriminated from within the noise burst is taken as a measure of temporal resolution threshold. Thresholds measured using this method are approximately 2-7msec (Formby, Morgan, Forrest, & Raney, 1992; Penner, 1977; Plomp, 1964; Shailer & Moore, 1983).

Auditory temporal resolution has been shown to vary according to the intensity of the stimuli (Corso, 1980), the *relative* intensities of the two stimuli (Formby et al., 1992; Gescheider, 1967), relative frequency content of the stimuli (Formby et al., 1992), and age (Gelfand, Porrazzo, & Silman, 1988). The nature of these relationships is qualitatively similar to within the visual system, i.e. temporal resolution increases with increasing stimulus intensity and declines with increasing age.

The auditory equivalent of the visual CFF is the Auditory Flutter Fusion threshold (AFF), the maximum temporal frequency of a fluttering sound which can be reliably discriminated from a continuous sound. In optimum conditions, this value is 1000Hz or more (Miller & Taylor, 1948; Symmes, Chapman, & Halstead, 1955). Although apparatus limitations precluded conclusive measurement of the upper limit of the AFF, it was estimated as being approximately 2000Hz (Miller & Taylor, 1948). Thus, as in the visual domain, temporal resolution appears to be slightly greater for streams rather than pairs of stimuli. AFF varies as a function of the sound-time fraction (the duration of noise expressed as a proportion of the total length of the stimulus stream, e.g. a sound-time ratio of 0.5 indicates that noise made up half the total length of the stimulus period), with a fraction of 0.5 resulting in highest AFF (Miller & Taylor, 1948; Symmes et al., 1955). Increasing signal intensity increases AFF (Besser, 1967), whilst ingestion of sedative drugs decreases it (Besser, Duncan, & Quilliam, 1966).

In the tactile modality, temporal resolution of two stimuli is possible when the stimuli are separated by approximately 3-18msec (Fucci & Petrosino, 1984; Fujisaki & Nishida, 2007a; Geldard & Sherrick, 1971; Gescheider, 1966, 1967, 1974; Petrosino & Fucci, 1989). All these values were obtained using physical taps to the skin surface; evidence suggests that thresholds are slightly higher (10-45msec) when using electrocutaneous stimulation (Hoshiyama, Kakigi, & Tamura, 2004; Lacruz, Artieda, Pastor, & Obeso, 1991; Rosner, 1961; Uttal, 1959). Vibrotactile gap detection thresholds appear to be greater than two-point temporal resolution thresholds at approximately 19-35msec (Formby et al., 1992). This implies that sustained vibrotactile stimulation may produce persistence of sensation which makes it difficult for observers to resolve small temporal gaps, even though transient ‘taps’ separated by similar gaps may be resolved with relative ease.

It appears that tactile temporal resolution thresholds are stable and repeatable over a one-year period (Petrosino & Fucci, 1985). Resolution is affected by the intensity of tactual stimulation (Gescheider, 1967; van Doren, Gescheider, & Verrillo, 1990), *relative* intensities of the two stimuli (Gescheider, 1966), and age (Hoshiyama et al., 2004; Petrosino & Fucci, 1989; van Doren et al., 1990); as in the visual and auditory systems, temporal resolution increases with increasing intensity, and reduces with increasing age.  It is likely that the decline in resolving power with age is related to the reported decline in perceived tactile intensity with age (Verrillo, 1982).  It is also likely that resolution varies with the size of the tactile stimulus on the skin surface, due to the effect of this variable on perceived intensity (Verrillo, 1963), although this remains speculatory.  Using electrocutaneous stimulation, Hoshiyama and colleagues demonstrated that temporal resolution varies between different parts of the body (Hoshiyama et al., 2004).  This variation is proposed to be related to differences in the somatotopic representation of different bodily regions in S1; however, differences in the electrical impedance of the skin at different regions are an equally plausible explanation, as the same current was used throughout the experiments (Hoshiyama et al., 2004).

In comparison with the visual and auditory domains, there appears to be an absence of literature investigating the flicker fusion thresholds for a stream of tactile stimuli.  This absence has historically been ascribed to the technical limitations of vibrotactile transduction apparatus  (Sherrick & Cholewiak, 1986).  However, based on the temporal resolution of the tactile system (as discussed above), it is likely that flicker fusion thresholds lie somewhere between visual and auditory values.

Cortical areas implicated as vital to tactile temporal resolution tasks (using electrocutaneous stimulation) include the pre-Supplementary Motor Area (pre-SMA) and anterior cingulate (Pastor, Day, Macaluso, Friston, & Frackowiak, 2004). Unsurprisingly, lesions to S1 cause substantial elevations in temporal two-point discrimination thresholds (Lacruz et al., 1991). It has also been shown that the basal ganglia and cerebellum are important in tactile temporal resolution (Lacruz et al., 1991). However, Pastor and colleagues found that these areas were activated during a spatial discrimination task as well a temporal two-point discrimination task (Pastor et al., 2004). The specificity of these areas for *temporally* resolving external events is therefore unclear. However, it is known that both structures are implicated in a range of timing tasks (Sections 2.1.4 and 2.1.6).

### 2.1.3 Judgments of temporal order and simultaneity

In a temporal order judgement (TOJ) task, observers are presented with two stimuli at a range of temporal offsets (Stimulus Onset Asynchrony (SOAs)) and are forced to choose which of the two appeared first (Section 3.3.1). Using this task, at least two measures of temporal order processing can be obtained. They are the Point of Subjective Simultaneity (PSS), or the *physical* temporal relationship between the stimuli which corresponds to *perceptual* simultaneity, and a measure of observers' sensitivity to asynchrony (i.e., the Just-Noticeable Difference (JND); Section 3.1.2). The two stimuli can be in either the same ('unimodal' or 'within-modality' TOJ) or different ('crossmodal' TOJ) sensory modalities.

In a unimodal TOJ task, the stimuli and methods used often give rise to concern. For example, in a visual TOJ task the stimuli are frequently differentiated by location (Jaskowski, 1992, 1993; Mitrani, Shekerdjiiski, & Yakimoff, 1986), so that the observer has to judge whether the right/top or left/bottom stimulus appeared first. The transient and sequential appearance of two lights at different spatial locations is likely to give rise to an apparent motion percept, such that the task is likely to resemble a motion direction discrimination judgment rather than a pure TOJ. The same concern may be applied to auditory TOJ studies where the stimuli are presented either to the right or left ear (Corso, 1980; Hirsh, 1959), or tactile TOJ where the stimuli are presented to different bodily regions (Shore, Gray, Spry, & Spence, 2005). Similarly, in a judgment of which one of two auditory stimuli differing in pitch was presented first (Hirsh & Sherrick, 1961; Kanabus, Szelag, Rojek, & Poppel, 2002), an observer is likely to use spectral cues (i.e., whether the stimulus pair appeared to ascend or descend in pitch) to assist their temporal judgment. Therefore, data derived from unimodal TOJ tasks are likely to measure other aspects of perception in addition to temporal order, and could be considered less valid as a measure of temporal perception than crossmodal TOJ data.

As both RTs (Section 2.1.1) and neural latencies (Sections 1.1.1 and 1.1.2) indicate that auditory stimuli are processed more rapidly than visual, we may expect that a visual stimulus must be presented slightly before an auditory stimulus in order for the two to be perceived as simultaneous. If conceptualised as a ‘race’ to the relevant cortical areas, the visual stimulus is slower than the auditory and therefore needs a ‘head start’ for the two to be perceived simultaneously. However, controversy exists within the body of work on this topic; some authors have found that a small *auditory* lead is required for the two stimuli to be perceived as simultaneous (Arnold, Johnston, & Nishida, 2005; Engel & Dougherty, 1971; Fujisaki, Shimojo, Kashino, & Nishida, 2004; Hamlin, 1895; Harrar & Harris, 2005, 2008; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Navarra et al., 2005; Neumann, Koch, Niepel, & Tappe, 1992; Rutschmann & Link, 1964; Smith, 1933; Sugita & Suzuki, 2003; Teatini, Farne, Verzella, & Berruecos, 1976), suggesting that simple RT cannot be used to predict PSS, whereas some authors find that a *visual* lead is required (Dinnerstein & Zlotogura, 1968; Exner, 1875; Hirsh & Sherrick, 1961; Jaskowski, Jaroszyk, & Hojan-Jezierska, 1990; Keetels & Vroomen, 2005; Kopinska & Harris, 2004; Lewald & Guski, 2004; Machulla, Di Luca, & Ernst, 2007; Spence, Baddeley, Zampini, James, & Shore, 2003; van Eijk, Kohlrausch, Juola, & van de Par, 2008; Vatakis, Bayliss, Zampini, & Spence, 2007; Zampini, Shore, & Spence, 2003a). Another study found that PSS did not systematically differ from zero, i.e. perceived simultaneity approximated to physical simultaneity (Heron, Whitaker, & McGraw, 2004). This controversy remains unresolved.

In the audiotactile (AT) pairing, a tactile lead appears to be required for perceptual simultaneity (Dinnerstein & Zlotogura, 1968; Harrar & Harris, 2008; Hirsh & Sherrick, 1961; Machulla et al., 2007; Navarra, Soto-Faraco, & Spence, 2007; Zampini et al., 2005a), as would be predicted from RT and neural latency data. In the visuotactile (VT) pairing, previous work indicates that a small visual lead is necessary for the two stimuli to be perceived as simultaneous (Dinnerstein & Zlotogura, 1968; Machulla et al., 2007; Poliakoff, Shore, Lowe, & Spence, 2006; Spence et al., 2003; Spence, Shore, & Klein, 2001b), as expected from the slower RT and neural latency to visual stimuli. In contrast, other VT studies have found that a *tactile* lead is necessary for perceptual simultaneity (Harrar & Harris, 2005, 2008; Hirsh & Sherrick, 1961; Occelli, Spence, & Zampini, 2008b). As in the AV domain, the reason(s) for this apparent dichotomy within the literature is currently unclear. Illustrative crossmodal TOJ results are shown as Figure 2.4.

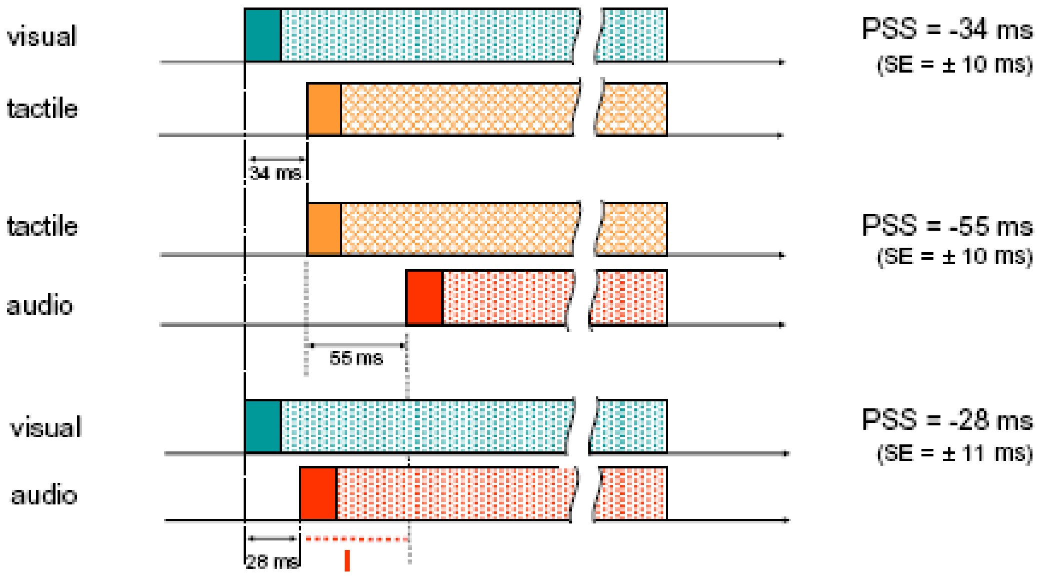


**Figure 2.4.** Mean PSS values (n = 16) obtained using a TOJ tasks in the VT, AT and AV pairings. Vision has to physically lead touch by 34msec for perceptual simultaneity, touch lead sound by 55msec, and vision lead sound by 28msec (Machulla et al., 2007).

An alternative to the TOJ task is the Simultaneity Judgment (SJ), sometimes referred to as the Synchronous-Asynchronous judgment (SAS). In this task, observers are presented with pairs of stimuli at a range of SOAs in an identical manner to the TOJ task. However, in the SJ task observers have to judge whether the stimuli are simultaneous or successive rather than ‘which came first'. The SOA at which an observer is most likely to respond that the two events are simultaneous is taken as the measure of PSS.

Crossmodal SJ results in the literature mainly pertain to the AV modality pairing. Using this task, it appears that observers require a *physical* lead of vision over sound for *perceptual* simultaneity (Stone et al., 2001; van Eijk et al., 2008; Zampini, Guest, Shore, & Spence, 2005b), although note that Smith (1933) found that an insignificant (2msec) lead of sound was required. However, considerable inter-observer variability has been reported in PSS values. For example, the 17 observers used in the study of Stone and colleagues displayed PSS values ranging from -21msec (a physical lead of sound over vision) to +150msec (a lead of vision over sound). Interestingly, PSS values were found to be highly repeatable for each observer (Stone et al., 2001). In the VT domain, average PSS appears to correspond to a physical lead of touch over vision (Vogels, 2004). To date, only Fujisaki and Nishida appear to have compared SJ data in the AV, AT and VT pairings (Fujisaki & Nishida, 2007a). This work aimed to measure the sensitivity of observers' SJs (Figure 2.7), and as such did not present a measure of PSS in any of the three modality pairings.

The main conclusion we may draw from examination of the PSS values derived from TOJ and SJ tasks (as reported above) is that the PSS appears to be observer-specific. Inter-observer differences in perceived timing have been known since the first publication of literature describing the so-called *personal equation* in the 19th century (Bessel, 1822). This concept was developed in astronomical observatories at a time when astronomers measured the transit time of stars with reference to an auditory timing device. Frequent differences in timing measurements recorded by different observers (famously, the Astronomer Royal at Greenwich Observatory fired his assistant for perceived ‘errors’ in his measurements in 1796) led Bessel to directly compare his judgments with that of another observer. This led to the concept of the personal equation as a time quantity which was added or subtracted to an observers’ timing judgment in order to compare it with the judgment of another observer (Bessel, 1822; Mollon & Perkins, 1996; Sanford, 1888a). These inter-observer differences have been suggested to reflect differences in the division of attention between vision and audition (Sanford, 1888b).

PSS has been demonstrated to vary as a function of selective attention to one of the stimuli in a TOJ task, an example of the *prior entry* effect (Frey, 1990; Shore, Spence, & Klein, 2001; Spence et al., 2001b; Stelmach & Herdman, 1991; Vibell, Klinge, Zampini, Spence, & Nobre, 2007; Zampini, Shore, & Spence, 2005c). The Law of Prior Entry states that an attended object is processed more rapidly than an unattended object (Titchener, 1908). Therefore, in a TOJ task with attention focused on one modality or location, we would expect that a stimulus presented in that modality/location would be perceived as occurring before a physically simultaneous stimulus presented in another modality/location. This would result in different PSS values being obtained, depending on whether attention was divided equally between modalities/locations (as is usual in TOJ tasks), or focussed on one modality/location (Shore et al., 2001; Spence et al., 2001b; Zampini et al., 2005c).

Given the temporal limitations of the peripheral transduction mechanisms in the three sensory modalities, at first consideration the Prior Entry effects discussed above appear improbable - for example, it is difficult to imagine transduction of auditory events at the cochlea occurring more rapidly via attentional modulation alone. However, recent evidence suggests that the *magnitude* of stimulus-evoked brain activity, rather than the *timing* of this activity, correlates with behavioural measures of perceived timing (McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2005; Noguchi & Kakigi, 2006). For example, a recent VT TOJ task demonstrated a prior entry effect which correlated with increased amplitude of early Event-Related Potential (ERP) components (Vibell et al., 2007). In contrast, latency differences of the same components were minimal, although latency shifts in later ERP components were somewhat larger (yet still significantly less than the size of the prior entry effect as demonstrated by behavioural data) (Vibell et al., 2007). This is consistent with prior entry being initially facilitated by an increase in amplitude of neural activity, and this increased activity being translated into reduced latency at higher cortical regions. Work examining visual RT to complex stimuli has also demonstrated a strong correlation between amplitude of neural activity (as measured with MEG) and RT, with greater neural activity being associated with faster RT (Amano et al., 2006).

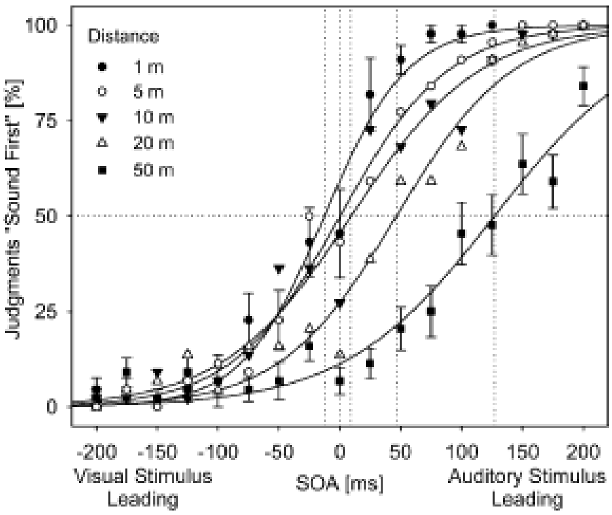


**Figure 2.5.** AV PSS measured at observer-source distances ranging from 1-50m. The physical temporal offset (SOA) corresponding to 50% sound-first responses on the y-axis is the PSS. With increasing distance, an increased physical lead of sound is required for perceived simultaneity (Lewald & Guski, 2004).

Peculiar to the AV domain is the effect of observer-source distance on relative timing judgments. As light travels significantly faster than sound in air (300,000km/sec vs. 0.343km/sec), the visual component of a distal event arrives at an observer before the auditory component. It has been claimed that when making AV temporal judgments, observers are able to compensate for these different propagation velocities in order to maintain a veridical perception of AV synchrony over a range of observer-source distances up to approximately 40m (Alais & Carlile, 2005; Engel & Dougherty, 1971; Kopinska & Harris, 2004; Sugita & Suzuki, 2003). Such a mechanism would have clear ecological benefits – AV stimuli originating from the same external event would always be perceived as simultaneous, despite the visual stimulus arriving at the observer before the auditory stimulus. However, other studies claim that no such compensation occurs (Arnold et al., 2005; Heron et al., 2007; Lewald & Guski, 2004; Stone et al., 2001). These studies maintain that with increasing distance observers require an increasing physical temporal lead of sound over vision for perceptual simultaneity; changes in AV PSS measured at varying distances thus approximate to that predicted by the speed of sound in air. Figure 2.5 illustrates changes in PSS as a function of observer-source distance in an AV TOJ task.

Differences in the experimental methods employed in these various studies may explain this apparent dichotomy within the literature. For example, all auditory stimuli in the study of Sugita and Suzuki (2003) were presented over headphones and contained no distance cues whatsoever. In the work of Alais and Carlile (2005), all visual and auditory stimuli were presented at a fixed distance of 57cm, with only the auditory stimuli containing any distance cues; these cues were simulated by manipulating the ratio of direct to reverberant energy of the stimuli (Figure 2.6). Finally, Engel and Dougherty (1971) and Sugita and Suzuki (2003) asked their observers to *imagine* that auditory and visual stimuli were co-localized in space at each distance tested. Such a cognitive strategy has been shown to produce PSS shifts consistent with neural compensation for the different speeds of sound and light in air (Arnold et al., 2005).

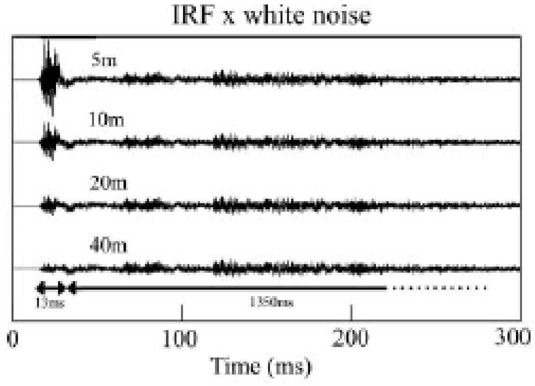


**Figure 2.6.** Waveforms of the auditory stimuli used in the study of Alais and Carlile (2005). Stimuli consisted of a 13msec ‘direct’ portion (which did not contain reverberant information) and a reverberant ‘tail’ 1,350msec long. Halving the amplitude of the direct portion of the waveform doubled the perceived observer-source distance by modulating the ratio of direct-reverberant energy. The impulse response function used to generate the reverberant ‘tail’ was recorded in the concert hall of Sydney Opera House (Alais & Carlile, 2005).

In marked contrast, those studies which concluded that observers were unable to compensate for distance in order to maintain perceived AV synchrony used co-localised auditory and visual stimuli without simulated distance cues. For example, the environments used included corridors (Arnold et al., 2005; Heron et al., 2007), a grassed outdoor area (Lewald & Guski, 2004), a large reverberant chamber (Heron et al., 2007) and an indoor laboratory (Stone et al., 2001). It must be noted that Kopinska and Harris (2004) used a similar, naturalistic paradigm and environment and found that observers' PSS was constant over distance; the reasons for this finding are unclear. Nevertheless, the weight of evidence suggests that humans are unable to compensate for the different propagation velocities of sound and light in order to maintain AV synchrony, at least within a natural environment. Such an absence of compensation is often observed in day-to-day life: for example, when watching a firework display we frequently see the fireworks explode before hearing the explosion.

It has also been reported that exposure to pairs of asynchronous AV stimuli (e.g., vision leading sound) modulates observers' perception of AV simultaneity. Specifically, PSS is found to shift in the direction of the adapting lag (Fujisaki et al., 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004); for example, following exposure to 'vision leads sound' stimulus pairs, a *physical* lead of vision over sound is required for *perceptual* simultaneity. A detailed review of the literature on this topic is presented in Sections 5.1.1 and 5.3.4.

The sensitivity of observers to changes in SOA has been shown to be approximately the same in the unimodal visual, auditory and tactile domains as well as the crossmodal combinations of these modalities, although unfortunately no quantitative analysis was presented to confirm this similarity (Hirsh & Sherrick, 1961). This suggests that the temporal resolution of the individual modalities is of little significance when judging the temporal order of stimuli (see Spence et al. (2001b) for a similar finding in the VT and unimodal V and T conditions). It is also consistent with the notion of a single neural mechanism being used for all TOJs. However, the unimodal data are subject to the potential methodological concerns already discussed (i.e., ‘temporal’ judgments which may additionally be influenced by other perceptual information such as apparent motion cues).

Observer sensitivity or Just-Noticeable Difference (JND; Section 3.3.1) has been suggested to vary as a result of the relative spatial locations of the stimuli in a crossmodal TOJ task; specifically, when the two stimuli are spatially co-localised, JNDs are higher than when the stimuli are spatially separated. This has been demonstrated in the AV (Keetels & Vroomen, 2005; Zampini et al., 2003a; Zampini, Shore, & Spence, 2003b) and VT (Spence et al., 2003; Spence et al., 2001b) pairings. This finding is usually considered to reflect observers using cues as to which location came first (rather than simply which modality came first) to aid their TOJ. However, work by Zampini and colleagues in the AV domain has suggested that this ‘spatial’ effect on JND is only evident when the different locations are either side of observers’ midline, resulting in different cerebral hemispheres being activated (Zampini et al., 2003b). Additionally, Fujisaki and Nishida, using a synchrony detection task with both pairs and streams of AV stimuli, found no effect of location on sensitivity to asynchrony (Fujisaki & Nishida, 2005, 2007b). In the AT pairing, the relative locations of stimuli have been shown not to affect JND (Zampini et al., 2005a). In summary, the evidence for relative stimulus location modulating JND in a TOJ task is mixed. It has been proposed that any such effects are specific to inexperienced observers, who presumably would benefit most from an additional spatial cue (Fujisaki & Nishida, 2005).

Using a SJ task, Fujisaki and Nishida examined the sensitivity of observers' timing judgments using AV, AT and VT stimulus pairs (Fujisaki & Nishida, 2007a). Performance was more accurate in the AT pairing (average SOA required for accurate performance = 37msec) than the VT (60msec) or AV (79msec) pairings. To a large extent, this finding is intuitive in that audition and touch are the two modalities with the highest temporal resolution (Section 2.1.2); it is therefore credible that this pairing offers greater temporal accuracy than either of those featuring vision[1]. In contrast to measures of visual and auditory unimodal temporal resolution (CFF and AFF, respectively), performance is greater for paired than streams of stimuli (Fujisaki & Nishida, 2007a). A graph illustrating the results of this study is shown as Figure 2.7.

Although measures of sensitivity to asynchrony derived from SJ data are likely to be dependent on observers' response criteria and thus potentially inaccurate (Section 3.3.3), the pattern of results agreed with previous work

---

[1] Note that sensitivity is unrelated to PSS: greater sensitivity does not imply that SJs in the AT pairing are more likely to be veridical.

which employed an asynchrony matching task (Sinex, 1978). Both of these studies also found that within-modality discrimination performance was significantly better than crossmodal (Fujisaki & Nishida, 2007a; Sinex, 1978). This difference with the finding of Hirsh and Sherrick (1961) may be due to the use of tasks other than TOJ to investigate temporal processing, and all these studies may also be confounded by the methodological issues inherent in unimodal relative timing judgments (see above).

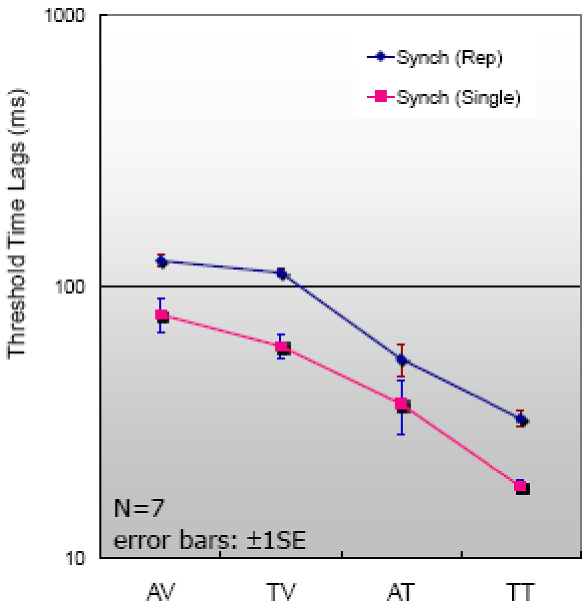


**Figure 2.7.** Graph showing the minimum SOA required for accurate synchronous/asynchronous judgments of pairs (pink squares) and streams (blue diamonds) of AV, AT, VT and unimodal tactile stimuli. Note that crossmodal resolution is highest in the AT pairing, but higher still in the unimodal tactile condition. Streams of stimuli yield higher thresholds than single pairs (Fujisaki & Nishida, 2007a).

The neural mechanism(s) underpinning crossmodal and within-modality TOJ and SJ is/are currently unclear, due to a paucity of studies examining this issue. However, evidence suggests that a cortical and subcortical network comprising the insula, cerebellum, inferior frontal gyrus, inferior parietal lobe, superior colliculus and posterior thalamus is responsible for detection of AV asynchrony in a SJ task (Bushara, Grafman, & Hallett, 2001). As both the superior colliculus (Stein et al., 2005) and insula (Hicks, Benedek, & Thurlow,

1988) contain cells responding to uni-, bi- and trimodal events in the visual, auditory and tactile modalities, it is possible that the neural connectivity exists for this network to also facilitate AT and VT SJs.

Given that SJ and TOJ may represent different measures of temporal processing (Section 3.3.3), it is feasible that an AV TOJ task may employ a slightly different neural network to that described by Bushara et al. (2001), although this is not currently known. It has even been proposed that there exist at least three different neural mechanisms mediating AV TOJ, which process temporal order according to stimulus characteristics (Fink, Ulbrich, Churan, & Wittmann, 2006). However, this result should be treated with caution due to the methods and stimuli used to investigate within-modality TOJ. Specifically, auditory stimuli were differentiated by pitch and by location, whilst the visual stimuli were also differentiated by location. As already stated, in these circumstances observers can use apparent motion cues or pitch cues (ascending vs. descending) to make their TOJ. The apparent intransitivity (Section 5.4) of multimodal PSS, at least using a TOJ task (Machulla et al., 2007), may also suggest that relative timing judgments employ different neural mechanisms in the three crossmodal pairings. However, no data examining this issue are currently available in the literature. Clearly, the neural correlates of multimodal TOJs and SJs require further investigation.

### 2.1.4 Perceived duration

As with the other measures of temporal processing already discussed, clear differences between the sensory modalities are evident when judging durations. In the first study directly comparing performance in the visual, auditory and tactile modalities, it was found that performance in an interval discrimination task is highest in the auditory modality, intermediate in the tactile modality, and poorest in the visual modality (Goodfellow, 1933). In other words, observers are able to detect smaller differences in duration between two unfilled intervals (defined by the elapsed time between two brief signals denoting the start and end of the interval; conversely, filled intervals are defined by a single signal of duration equal to the interval) when the intervals are marked by auditory events than by either tactile or visual events; the auditory judgments are more precise. A similar pattern has been observed when using filled visual and auditory (but unfilled tactile) intervals (Westheimer, 1999). The superior accuracy of audition relative to vision is also clear from a number of other studies (Grondin & Rousseau, 1991; Rousseau, Poirier, & Lemyre, 1983; Tanner, Patton, & Atkinson, 1965).

When discriminations of duration are made *between* (rather than within, as discussed above) modalities, accuracy is significantly reduced relative to within-modality discriminations (Grondin & Rousseau, 1991; Rousseau et al., 1983; Westheimer, 1999). This result has obvious parallels with the asynchrony detection data discussed in Section 2.1.3 (Fujisaki & Nishida, 2007a; Sinex, 1978), part of which is illustrated as Figure 2.7. It has been suggested that switching attention across modalities may make the task more difficult and therefore reduce the accuracy of observers' duration judgments (Mauk & Buonomano, 2004), a hypothesis which is credible given that shifting attention between modalities is known to impair performance in other temporal tasks (Spence et al., 2001a).

At present, it appears that no researchers have measured relationships between perceived visual, auditory and tactile durations within the same study, and that the majority of studies investigating perceived duration have compared visual and auditory intervals. These studies are unanimous in finding that auditory intervals are perceived as being longer than visual intervals of the same objective duration; this pattern is true of both filled and unfilled intervals (Behar & Bevan, 1961; Goldstone & Goldfarb, 1963; N'Diaye, Ragot, Garnero, & Pouthas, 2004; Walker & Scott, 1981; Wearden, Edwards, Fakhri, & Percival, 1998; Wearden, Todd, & Jones, 2006). Moreover, when visual and auditory stimuli of the same physical duration are simultaneously presented, the perceived duration of the stimulus combination is the same as that of an auditory stimulus of the same duration presented alone (Walker & Scott, 1981). This suggests that greater perceptual weight is given to auditory input than visual when making duration judgments. Given the superior accuracy of auditory duration judgments, this is credible and likely to be in keeping with the concept of *modality appropriateness* (Welch & Warren, 1980) discussed in Section 2.1.5.

The only study investigating visual and tactile (unfilled) intervals similarly found that tactile intervals were perceived as longer than visual by approximately 9% (van Erp & Werkhoven, 2004). When auditory and tactile intervals (filled) were compared, they were found not to differ significantly in perceived duration (Ehrensing & Lhamon, 1966; Hawkes, Deardorff, & Ray, 1977). Therefore, it appears that both auditory and tactile durations are perceived as longer than visual. In the AV context, this has been suggested to be due to differences in temporal resolution in the visual and auditory modalities (Ivry & Schlerf, 2008) – but if this were the case, we would also expect auditory intervals to be perceived as longer than tactile, which appears not to be the case (Ehrensing & Lhamon, 1966; Hawkes et al., 1977). Alternatively, Wearden and colleagues have proposed that an auditory timing mechanism (or ‘clock’) runs faster than its visual counterpart (Wearden et al., 1998).

Perceived duration may be modulated by a wide range of stimulus and task manipulations. For example, infrequent or unexpected stimuli are perceived as being longer than frequent or expected stimuli of the same objective duration (Pariyadath & Eagleman, 2007; Tse, Intriligator, Rivest, & Cavanagh, 2004; Ulrich, Nitschke, & Rammsayer, 2006), and the first of a stream of identical stimuli appears longer in duration than those stimuli following it (Kanai & Watanabe, 2006; Pariyadath & Eagleman, 2007; Rose & Summers, 1995). The effect of this variable on perceived duration is shown as Figure 2.8. Other factors shown to affect perceived visual duration are non-temporal stimulus dimensions such as size and numerosity (Xuan, Zhang, He, & Chen, 2007), stimulus visibility (Terao, Watanabe, Yagi, & Nishida, 2008), flickering visual stimuli (Kanai, Paffen, Hogendoorn, & Verstraten, 2006), saccadic eye movements (Morrone, Ross, & Burr, 2005), and deployment of attentional resources (Chaston & Kingstone, 2004).

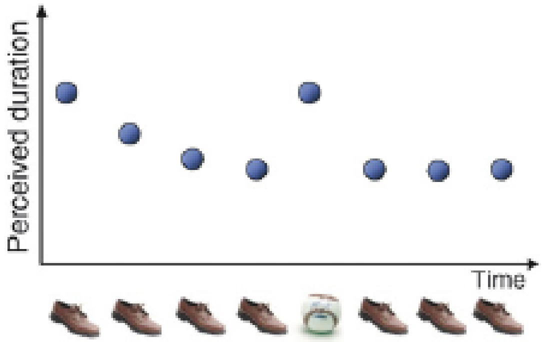


**Figure 2.8.** Effect of stimulus predictability and presentation order on visual perceived duration. The first stimulus is perceived as longer than the identical stimuli which follow it, and an 'oddball' stimulus (in this example, a baseball) is perceived as longer than heterogeneous stimuli of identical duration (Eagleman, 2008).

The concept of duration is vital to our understanding of temporal processing, as it is used to measure the passage of time. For this reason, the neural structures and mechanisms of duration are of particular interest. A study of this area is complicated by the fact that temporal processing in humans occurs on at least four different scales: microseconds, milliseconds, seconds, and circadian rhythms (Buonomano & Karmarkar, 2002; Mauk & Buonomano, 2004). For the purposes of this brief overview, I will focus on temporal processing in the millisecond range as this is considered to be relatively free of the confounding influence of higher-level, attentive factors (Buonomano & Karmarkar, 2002; Lewis & Miall, 2003), and such durations are typically too short for observers to employ a counting strategy to aid their temporal judgments.

It is likely that the cerebellum is an important neural structure in temporal processing on the millisecond scale.  This has been shown with duration discrimination tasks (Ivry & Keele, 1989; Jueptner et al., 1995; Tregellas, Davalos, & Rojas, 2006), temporal reproduction tasks (Bueti, Walsh, Frith, & Rees, 2008c; Jahanshahi, Jones, Dirnberger, & Frith, 2006), and time estimation tasks (Bueti et al., 2008c; Smith, Taylor, Lidzba, & Rubia, 2003). Despite the likely importance of the cerebellum, a number of mechanisms and structures mediating temporal processing are likely to be distributed throughout the brain, as evidenced by a range of behavioural and neuroimaging studies.

It appears that there are multiple mechanisms for processing of specific time intervals in the hundreds of milliseconds range (Karmarkar & Buonomano, 2003; Nagarajan, Blake, Wright, Byl, & Merzenich, 1998).  These mechanisms are not selective for sensory modality (Nagarajan et al., 1998), auditory pitch (Karmarkar & Buonomano, 2003) or visual hemifield (Westheimer, 1999).  For example, observers trained to discriminate auditory intervals of approximately 100msec demarcated by 1kHz tones performed with similar accuracy when the intervals are demarcated by 3.75kHz tones, but discrimination thresholds for 200msec intervals showed no effect of training (Karmarkar & Buonomano, 2003).  It has also been shown that training on a duration discrimination (sensory) task improves performance when reproducing the trained duration using buttonpresses (Meegan, Aslin, & Jacobs, 2000).  The lack of specificity for task type (sensory vs. motor), low-level stimulus features and sensory modality suggest that these timing mechanisms operate at a late stage of neural processing, beyond modality-specific sensory cortices.

In contrast, other studies have produced data consistent with low-level cortical or subcortical timing mechanisms. Judgments of visual duration can be modulated in a spatially-specific manner following adaptation to flickering or oscillatory stimuli. Specifically, adaptation to 20Hz flicker reduces the perceived duration of stimuli presented at that location relative to unadapted locations (Johnston, Arnold, & Nishida, 2006). Preliminary work also suggests that this spatially-specific effect can also be observed in the tactile domain (Watanabe, Amemiya, Nishida, & Johnston, 2008). Modality-specific auditory (Bueti, van Dongen, & Walsh, 2008b) and visual (Bueti, Bahrami, & Walsh, 2008a) timing mechanisms have been recorded in the superior temporal gyrus (the brain region containing auditory cortex) and V5/MT, respectively, with the use of Transcranial Magnetic Stimulation (TMS). Finally, visual duration information was shown to be available to both cerebral hemispheres in a split-brain patient, indicating subcortical processing and transfer of temporal information (Handy, Gazzaniga, & Ivry, 2003; Marzi, 2004).

It has also been suggested that certain neural timing mechanisms are localised within the right cerebral hemisphere (Figure 2.9). For example, Harrington and colleagues found significantly elevated duration discrimination thresholds specific to patients with cortical lesions to the right hemisphere (Harrington, Haaland, & Knight, 1998). The cortical regions implicated were prefrontal, premotor and inferior parietal cortex (Harrington et al., 1998).

Similar results have been demonstrated in other studies (Funnell, Corballis, & Gazzaniga, 2003; Handy et al., 2003; Kagerer, Wittmann, Szelag, & Steinbüchel, 2002; Smith et al., 2003). However, it is also thought that there is a right hemisphere processing advantage for visual non-temporal information as well as temporal (Corballis, Funnell, & Gazzaniga, 2002; Handy et al., 2003), although this is not the case for audition (Harrington et al., 1998). Whether this hemispheric processing advantage is purely temporal is therefore open to debate; however, it is clear that differences in temporal information processing are evident between the two cerebral hemispheres (Grondin & Girard, 2005).

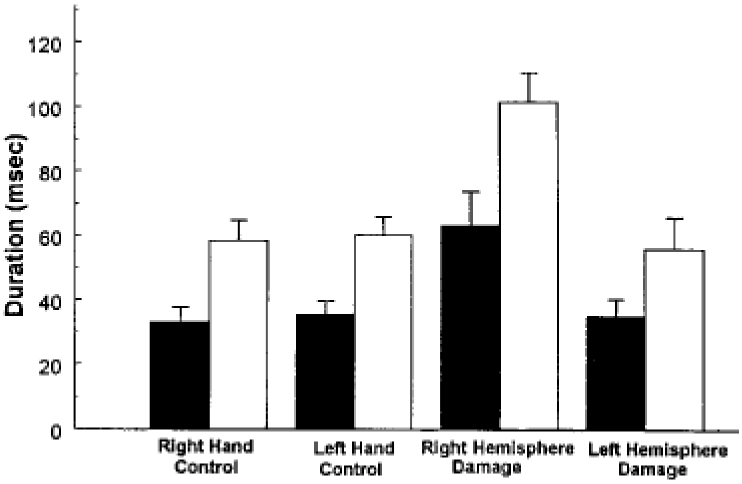


**Figure 2.9.** Average unfilled auditory duration discrimination thresholds (in msec) for normal control subjects and subjects with damage to the right and left cerebral hemispheres. Filled bars show thresholds for a standard duration of 300msec, unfilled bars for a standard of 600msec. Patients with right hemisphere lesions responded with their right hands, as did the subjects in the right hand control condition; the reverse was true for left hemisphere lesion patients and subjects in the left hand control condition. Discrimination thresholds are significantly elevated in those patients with right hemisphere damage (Harrington et al., 1998).

In order to reconcile these seemingly disparate findings, the only conclusion to be drawn is that temporal processing in humans is accomplished by a plurality of mechanisms and in a variety of cortical and subcortical structures. This is consistent with numerous distributed timing mechanisms, and rules out the existence of a single centralised timing mechanism within the brain (Buonomano & Karmarkar, 2002). However, it is also consistent with a recent proposal that temporal information is encoded by changes in the state of cortical networks over time, rather than dedicated timing mechanisms *per se* (Karmarkar & Buonomano, 2007). In summary, the precise mechanisms of timing within the brain are still under investigation.

### 2.1.5 Temporal rate perception

Just as differences in perceived duration arise between the sensory modalities, it has also been shown that perception of temporal rate differs between vision and audition. Specifically, fluttering auditory sequences are perceived as alternating more slowly than visual sequences flickering at the same rate, at least within the rate range 4-10Hz (Welch, DuttonHurt, & Warren, 1986). As with the differences in visual and auditory perceived duration (Section 2.1.4), this is consistent with the hypothesis of Wearden and colleagues that a timing mechanism or ‘pacemaker’ serving audition runs faster than that serving vision (Wearden et al., 1998).

This difference in perceived flicker/flutter rate in the visual and auditory modalities appears to influence observers’ judgments of temporal rate. Previous work by Knox has shown that observers judge visual flicker to be reduced in frequency when presented concurrently with fluttering auditory stimuli (Knox, 1945). When adjusting the rate of fluttering auditory stimuli to match the perceived flicker rate of a visual standard it was found that the perceived rate of the standard changed as the auditory comparison was adjusted, despite remaining objectively constant (Gebhard & Mowbray, 1959). This was investigated further by Shipley, who employed a task in which the flutter rate of an auditory stimulus was adjusted whilst a simultaneously presented visual stream remained constant; observers were required to judge whether the two were synchronous or asynchronous (Shipley, 1964). When both stimuli initially flickered/fluttered at a rate of 10Hz, the auditory rate could be varied between 7 and 22Hz without any apparent loss of synchrony of the two streams (Shipley, 1964). This strong influence of auditory rate on perceived visual rate is referred to as *auditory driving* (Welch & Warren, 1986).

Auditory driving has been shown to be unidirectional, i.e. visual stimuli appear to have minimal effect on perceived auditory rate (Recanzone, 2003; Welch et al., 1986), and to vary in effect magnitude with different temporal rates (Shipley, 1964). Importantly, it still occurs when a method of constant stimuli in conjunction with a forced-choice task (Section 3.1.2) is employed (Recanzone, 2003) rather than the method of adjustment (Gebhard & Mowbray, 1959; Myers, Cotton, & Hilp, 1981), a method more open to the influence of observers’ response biases and criterion (Sections 3.2.1 and 3.2.4).

Results of studies demonstrating auditory driving are in agreement with the hypothesis of *modality appropriateness* (Welch & Warren, 1980), which states that the observers allocate greater perceptual weight to the sensory modality likely to yield the most accurate information about the task in hand. As auditory temporal acuity is significantly greater than visual (Section 2.1.2), observers are biased toward this modality when making temporal judgements. The reverse is true in spatial tasks, hence the well-known ventriloquist illusion where the perceived location of a sound is biased toward the physical location of a temporally proximal visual event (Choe, Welch, Guilford, & Juola, 1975; Thomas, 1941). However, more recent work has suggested that this allocation of perceptual weight is not done in a mandatory fashion (e.g., automatically allocating processing resources to audition on all temporal tasks), but rather according to the reliability of the stimuli employed. By adjusting the modulation depth of fluttering auditory stimuli to a point where rate discrimination thresholds exceeded those in the visual modality (i.e., by degrading the reliability of the auditory stimuli relative to the visual stimuli), visual driving of auditory rate perception was elicited (Roach et al., 2006). Therefore, it appears that the unidirectional auditory driving effects reported previously (Recanzone, 2003; Welch et al., 1986) reflect the finer temporal resolution of auditory processing under most circumstances, but that the perceptual system is able to reverse this perceptual bias towards audition when auditory stimuli become less reliable than visual.

### 2.1.6 Sensorimotor timing

Although the literature reviewed thus far has pertained to purely sensory aspects of time perception, there is considerable evidence that observers' motor actions can profoundly affect time perception. Possibly the best-known example of this is the stopped clock illusion, where upon making an eye movement to a clock face the second hand appears to briefly pause before continuing at a normal pace. This was first empirically investigated by Yarrow and colleagues, who termed the effect *chronostasis* (Yarrow, Haggard, Heal, Brown, & Rothwell, 2001). The magnitude of this illusory slowing of subjective visual time was found to be dependent on the size of the saccadic eye movement (Yarrow et al., 2001), and subsequent work has demonstrated similar effects with motor actions other than saccades (Park, Schlag-Rey, & Schlag, 2003). Analogous effects have also been demonstrated in the auditory (Hodinott-Hill, Thilo, Cowey, & Walsh, 2002) and tactile (Yarrow & Rothwell, 2003) domains, following endogenous attentional shifts (Hodinott-Hill et al., 2002) and voluntary arm movements (Yarrow & Rothwell, 2003). This clearly indicates that the process of making voluntary actions may give rise to temporal misperceptions in the visual, auditory and tactile modalities. An example of this temporal misperception is the phenomenon of *intentional binding*.

Haggard and colleagues have demonstrated that the perceived timing of a sensory event can be modulated by the motor action which caused it.

Specifically, a motor action and subsequent auditory tone were shifted toward each other in perceived time, an effect the authors termed intentional binding (Haggard, Clark, & Kalogeras, 2002). This illusion appears to be dependent on the temporal delay between action and auditory event, in that the amount of temporal recalibration decays with increasing delay between action and event (Haggard et al., 2002). This has been interpreted as being consistent with the idea that the causal link between action and event is crucial to intentional binding (Eagleman & Holcombe, 2002), as with increasing time between an action and event the causal link between the two is reduced (Hume, 1748). Subsequent work demonstrated similar temporal attraction between motor actions and consequent tactile events (Tsakiris & Haggard, 2003). The importance of the causal link between action and event (or the *intention* to cause an effect) was confirmed by an absence of intentional binding when the movement was involuntary (Haggard et al., 2002; Tsakiris & Haggard, 2003).

A potential concern with the studies discussed above is the use of the rotating-clock method devised by Libet and colleagues (Libet, Gleason, Wright, & Pearl, 1983). The validity of this method has been subject to extensive debate, discussed in Section 3.3.4. However, more recent work has employed an interval estimation task (a relatively uncontroversial method of investigating perceived time) to demonstrate intentional binding between voluntary actions and visual, auditory and tactile events (Engbert, Wohlschläger, & Haggard, 2008). This suggests that the intentional binding effect is robust and not merely an artefact of the experimental paradigm used.

Although the authors do not speculate as to the neural structures responsible, the subsequent finding that schizophrenic subjects experience a greater magnitude of intentional binding than normals (Haggard, Martin, Taylor-Clarke, Jeannerod, & Franck, 2003) may implicate components of the dopaminergic pathway. The basal ganglia and medial forebrain are suggested as possible candidates due to their involvement in associating actions with their consequences in animals (Haggard et al., 2003). In schizophrenics, these dopaminergic circuits are known to be overactive (Meyer-Lindenberg et al., 2002), which would correlate with excessive intentional binding in this population. Finally, the basal ganglia have been implicated as being important in a range of timing tasks (Bueti et al., 2008c; Handy et al., 2003; Jahanshahi et al., 2006; Rao, Mayer, & Harrington, 2001).

Currently unknown is the relationship between intentional binding-like effects and an illusory reversal of perceived temporal order of action and effect first described by Cunningham et al. and quantified by Stetson et al. (Cunningham, Billock, & Tsou, 2001; Stetson, Cui, Montague, & Eagleman, 2006). These studies demonstrated that when observers are accustomed to a delay between their voluntary action and its subsequent sensory consequence, the same sensory event presented at a *reduced* delay appears to precede the motor action which caused it. The neural regions activated during this illusion are anterior cingulate cortex and medial frontal cortex, as seen in Figure 2.10 (Stetson et al., 2006). As these brain regions are implicated in conflict monitoring, the authors propose that the brain contains multiple representations of perceived temporal order, and the illusion causes competition between the competing representations (Stetson et al., 2006). This work is discussed in more detail in Section 7.1.1.

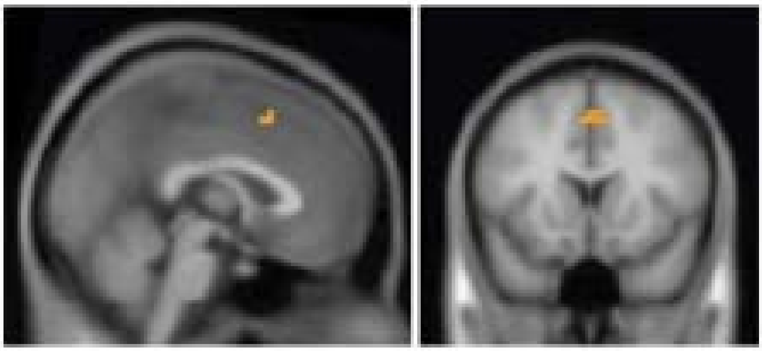


**Figure 2.10.** fMRI data showing brain regions activated during an illusory reversal of perceived temporal order. The activated voxels correspond to anterior cingulate cortex and medial frontal cortex (Stetson et al., 2006).

It has also been demonstrated that the magnitude of the flash-lag effect is reduced when the flash is perceived as being caused by a motor action (Lopez-Moliner & Linares, 2006). The flash-lag effect is an illusion whereby a stationary stimulus presented physically aligned with a moving bar appears to lag behind the bar (Figure 2.11). As in intentional binding, the authors found that a causal link between action and sensory consequence is essential for this reduction to occur, and additionally speculated that the causal link reduced the detection time of the flash (Lopez-Moliner & Linares, 2006).

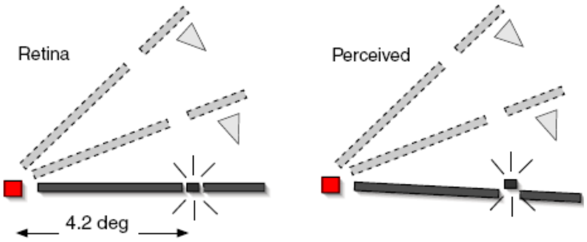


**Figure 2.11.** Example of the flash-lag effect. A stationary stimulus flashed in physical alignment with a moving stimulus appears to lag behind it (Lopez-Moliner & Linares, 2006).

Another example of timing errors associated with voluntary actions can be found in the pacing/tapping literature. In a typical pacing or tapping experiment, observers have to synchronise a series of voluntary movements (typically a tap of the finger) with a sensory event referred to as a ‘pacing signal’ (most commonly presented in the auditory modality). Task performance is usually measured by two parameters: the time difference between the sensory signal and observers’ taps (i.e., the temporal bias associated with synchronisation), and the timing variability associated with the taps.

The typical finding in such experiments is that perceptual synchronisation between an observers’ tap and the pacing signal is achieved by tapping prior to the pacing signal – in other words, observers make a systematic anticipatory error. This error is known as a *negative asynchrony* (Aschersleben, Stenneken, Cole, & Prinz, 2002). Typical magnitudes of the negative asynchrony when synchronising to an auditory pacing signal range from 20-80msec, with a large degree of inter-observer variability (Aschersleben, 2002). The magnitude of negative asynchrony appears to vary between observers, as is common with psychophysical measures of performance. Observers with some degree of musical proficiency are frequently found to exhibit a reduced degree of negative asynchrony relative to untrained subjects (Aschersleben, 1994; Repp, 1999), with the most highly-trained (e.g., concert-grade pianists) occasionally being able to synchronise exactly to an auditory signal (Repp, 1999). When comparing synchronisation to pacing signals presented in different modalities, data suggest that the negative asynchrony is largest when synchronizing to an auditory signal and smallest to a visual signal (Dunlap, 1910; Kolers & Brewster, 1985; Repp & Penel, 2002), with tactile signals giving rise to intermediate errors (Kolers & Brewster, 1985). Given that the timecourse of sensory transduction is slowest in the visual modality and fastest in the auditory modality, this pattern of results could be considered surprising.

The reason(s) that the negative asynchrony occurs is currently unclear. Two of the most influential explanations are the nerve-conduction hypothesis, also known as the Paillard-Fraisse hypothesis (Fraisse, 1980; Paillard, 1949), and the sensory accumulator model (Aschersleben, Gehrke, & Prinz, 2004). The nerve conduction hypothesis ascribes negative asynchronies to latency differences in the different sensory modalities, whilst the sensory accumulator model emphasises processes within the brain. As recently reviewed, neither of these hypotheses can completely account for all of the experimental findings associated with this body of literature and the mechanisms underpinning the negative asynchrony are still subject to debate (Repp, 2005).

The variability of the timing of observers' taps in a pacing task (i.e., the variability of the negative asynchrony) differs between the sensory modalities. Taps perceptually synchronised with an auditory event are less variable than those synchronised with a visual event (Kolers & Brewster, 1985; Repp & Penel, 2002, 2004), with those to a tactile signal exhibiting intermediate variability (Kolers & Brewster, 1985). As with negative asynchronies, variability is lowest with musically trained observers (Repp & Penel, 2002). It also decreases with decreasing intertap intervals (Semjen, Schulze, & Vorberg, 2000).

Another parameter investigated in pacing studies is the temporal limits of synchronisation, i.e. the maximum and minimum rate of event presentation to which observers can synchronise their actions. Observers can synchronise their taps to auditory events at presentation frequencies of approximately 8-10Hz, compared with only 2.5Hz in the visual domain (Repp, 2006); again, auditory performance appears to be superior to visual. In contrast, the lower limit of synchronisation is the same in both modalities at approximately 0.6Hz (Repp, 2006). Although no data pertaining to the tactile modality is currently available, it is likely (based on the negative asynchrony and variability results discussed above) that the upper limit of synchronisation is intermediate to that in the visual and auditory modalities.

Overall, characteristics of the negative asynchrony in the different modalities can be summarised as follows: observers are most precise (least variable) but most biased (least veridical) in the auditory condition, and least precise and least biased in the visual condition. Results in the tactile condition appear to be intermediate to the auditory and visual conditions.

# Chapter 3

## 3.1 Methodology

In this chapter, an overview of the common methods used in psychophysical measurement is presented, together with an introduction to some of the most important concepts to be considered when quantifying perceptual experience.

### 3.1.1 Basic principles of signal detection theory

The concept of signal detection theory (also known simply as detection theory or SDT) is an attempt to understand how organisms make decisions in conditions of uncertainty. This uncertainty is the result of the presence of 'noise' within the perceptual system. Noise is always present to some degree, and can be either internal or external (to the observer) in origin.

Internal noise can be caused by moment-to-moment variations in neural firing rates, which can be independent of the target stimulus (Corliss & Norton, 2002). All neurons occasionally fire without any external stimulation, creating a baseline level of neural activity or noise. This internal noise gives rise to observer uncertainty during the process of making a decision during a psychophysical task.

External sources of noise mainly relate to the presentation of the stimulus itself. For example, a light presented at a specific brightness level may be presented instead at a slightly different level (Corliss & Norton, 2002). This could be apparent as a slight variation in the number of photons emitted on each trial; in an auditory context, it may be apparent as minute trial-to-trial variations in intensity or spectral content. Therefore, the observer may be judging comparable but very slightly different events on different trials. Alternatively, noise can be added deliberately to the stimulus by the experimenter.

It is important to bear in mind that *all* sensory experience (visual, auditory, tactile, olfactory or gustatory; real or imagined) is coded by firing of neurons at some level of the central nervous system. As already stated, a variable level of neural activity is constantly present even in the absence of any stimulus. In order to (e.g.) detect the presence of a dim light, an observer must therefore discriminate the neural activity caused by the presence of the stimulus combined with background neural activity (signal + noise) from the background neural activity alone (noise) (Rose, 2006). This is illustrated in Figure 3.1.

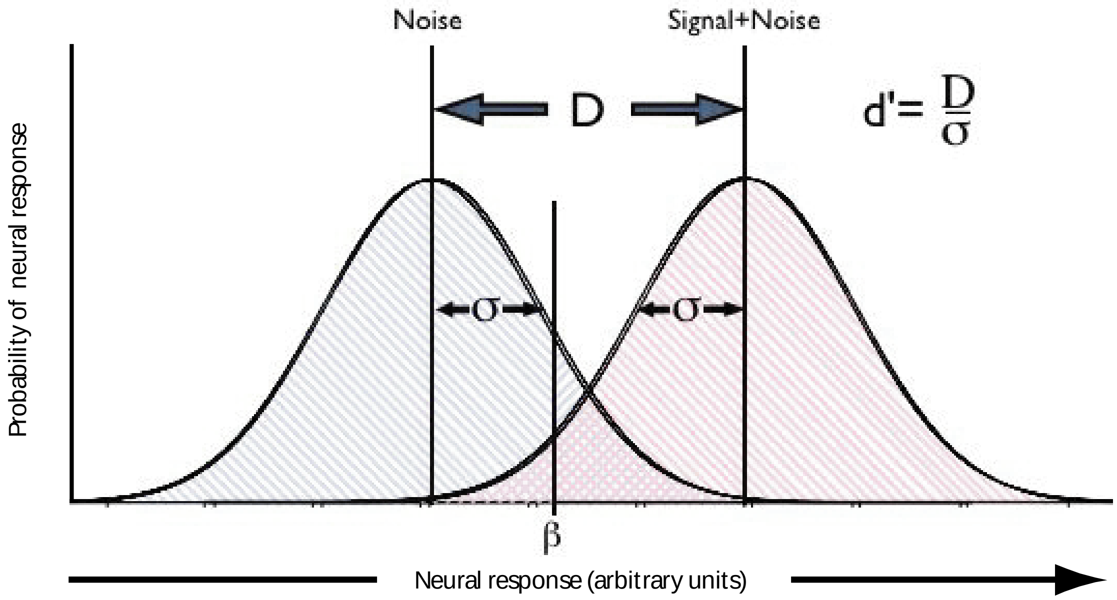


**Figure 3.1.** Diagram showing the hypothetical probability distributions of neural responses in a psychophysical task for 'Noise' (stimulus absent, background neural activity ('noise') present) and 'Signal + Noise' (stimulus and noise present) situations. The separation between the peaks of the distributions is given by D. The greater the separation between the peaks, the easier the signal is to discriminate from the noise (hence, the more likely an observer is to report the presence of the stimulus). The standard deviation of the Gaussian probability distributions is given by the symbol σ. Smaller values of σ equate to a less variable amount of noise in the neural system, and therefore σ is also a critical variable in stimulus detectability. This can be seen in the measure of discriminability d', which is given by dividing D by σ (details in text). To decide whether the stimulus is present or not, the observer must set some level of neural activity above which they respond that the stimulus is present and below which they respond that the stimulus is absent; this benchmark is referred to as the criterion adopted by the observer. In this figure, the criterion is shown by the symbol β on the x-axis. Neuronal responses falling below this level will cause the observer to respond 'absent', responses above this value will lead to a response of 'present'. Figure adapted from www.csic.cornell.edu/201/signal_detection.

From inspection of Figure 3.1 and accompanying text, it can be seen that the detectability of the stimulus is critically dependent on the separation of the noise (N) and signal + noise (S+N) distributions. The separation of the distributions is primarily affected by the intensity of the stimulus; for example, a loud tone will cause a greater increase in neuronal response than a quieter tone. Therefore, the S+N distribution in response to a bright light will be positioned further to the right of Figure 3.1 than the distribution corresponding to a dimmer light. The greater the distance between the peaks of the N and S+N distributions, the easier for observers to detect the signal (i.e., to discriminate signal from noise) (Rose, 2006).

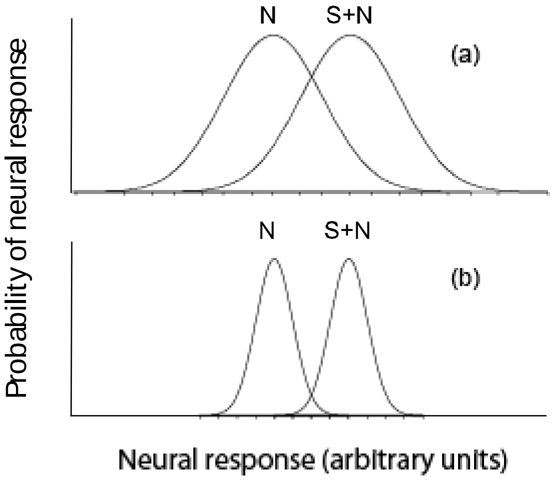


**Figure 3.2(a & b).** Examples of different amounts of noise variance (different values of σ), and corresponding effects on the appearance of the N and S+N probability distributions. In Figure 3.2(a), a relatively large amount of noise variance is present. The N and S+N distributions are correspondingly wide, with a large amount of overlap between the functions. In contrast, the probability distributions in Figure 3.2(b) are relatively narrow and with less overlap, reflecting a less variable amount of noise within the system and thus a smaller σ. Because of this reduced overlap of the distributions, the signal is easier to discriminate from the noise than in Figure 3.2(a). Note that the peaks of the N and S+N distributions are at the same location in both parts of the figure, indicating that the neural responses are to stimuli of equal objective intensity. Adapted from http://www.cns.nyu.edu/%7edavid/handouts/sdt-advanced.pdf.

Of similar importance in discrimination of the stimulus from the underlying noise is the standard deviation of the N and S+N distributions, usually denoted by the symbol σ. A low level of variability in the amount of noise in the neural system (i.e., a low value of σ) would mean that the probability distributions are relatively narrow in shape. With increasing σ, the functions would become wider in shape and spread over a correspondingly wider range of neural responses. In simple terms, the greater the overlap of the N and S+N distributions, the more difficult it is for observers to reliably discriminate the stimulus from the underlying noise. This is illustrated in Figure 3.2(a & b).

A variable amount of noise is inevitable, due to the random neural firing discussed previously. The greater the amount of noise present on any given trial, the more difficult the observers' decision. This increase in task difficulty is because of the close relationship between the standard deviation and separation of the probability distributions, and detectability. This can be expressed as detectability = separation/spread or, in the same notation as Figure 3.1,

$$d' = \frac{D}{\sigma}$$

The criterion (β) used by observers to judge the event of interest has a strong influence on the accuracy (or otherwise) of their judgments of the event of interest. In Figure 3.1, the observer has set their criterion at the point β on the x-axis. This means that any neuronal responses falling below this level will cause the observer to respond that the stimulus was absent, whilst any responses greater than this level will lead to a response that the stimulus was present. Because of the overlap between the N and S+N distribution functions, there are four possible outcomes on each trial: a hit, miss, false alarm or correct rejection. The relation of these outcomes to the probability distributions is shown in Figure 3.3(a & b).

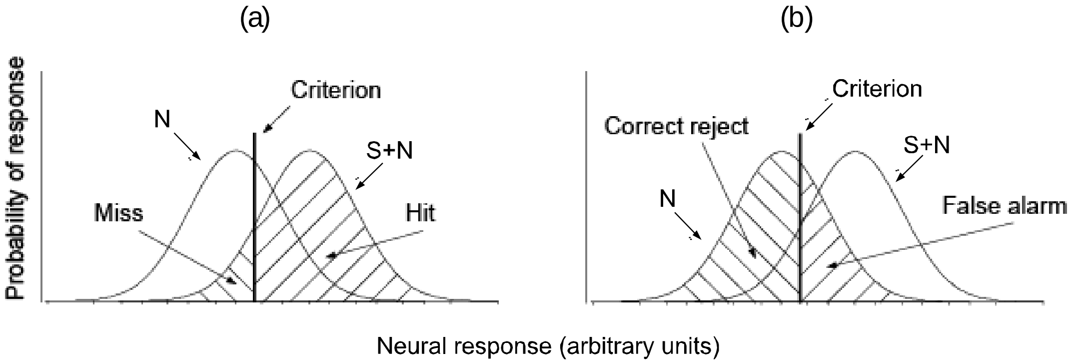


**Figure 3.3(a & b).** Diagrams showing the four possible outcomes of an observers' response during a psychophysical detection experiment. If the neural activity on a given trial exceeds that corresponding to criterion the observer will respond that a stimulus was present. If a stimulus was actually present, the trial is referred to as a hit; if no stimulus was present, the trial is known as a false positive. If the neural activity on any trial is less than that corresponding to criterion, the observer will respond that no stimulus was presented. If no stimulus was actually presented, the trial is referred to as a correct rejection; if a stimulus was present, the trial is known as a miss. Adapted from http://www.cns.nyu.edu/%7edavid/handouts/sdt-advanced.pdf.

From inspection of Figure 3.3(a & b) and accompanying text, it can be seen that observers' response criterion has a profound influence on the outcome of the experiment. The response criterion varies between observers and may vary within individual observers during the course of the experiment. Criterion can also be manipulated by the experimenter (Corliss & Norton, 2002). If hits were rewarded, for example, it would be expected that observers would lower their criterion (i.e., β would move to the left in Figure 3.1) to maximise the number of hits. This would also have the effect of increasing the number of false alarms, however. This may be acceptable if the consequences of a miss are severe (for example, if the observer were performing a task such as studying mammograms for evidence of abnormality), but is unlikely to be acceptable in other circumstances. Alternatively, should false alarms be penalised in some way, observers would be expected to raise their criterion (i.e., β would move to the right in Figure 3.1) in order to minimise the number of false alarms (Corliss & Norton, 2002). This would also raise the number of misses, however. This may be acceptable if a miss would result in relatively trivial consequences compared to the resources used in processing a large number of false alarms. The effects of criterion changes on the proportion of hits and false alarm responses are illustrated as Figure 3.4.

Importantly, there is no ‘perfect’ criterion that will completely eliminate incorrect responses. Because the noise and signal + noise probability distributions are normally distributed, the functions will always overlap to an extent which makes some proportion of errors inevitable (e.g. Figure 3.4, top panel). Lower values of σ (i.e., narrower noise variance distributions) and increased D values (i.e., increased distance between the distributions, corresponding to a more intense stimulus presentation) will give rise to fewer misses and false alarms, but will never eliminate them completely.

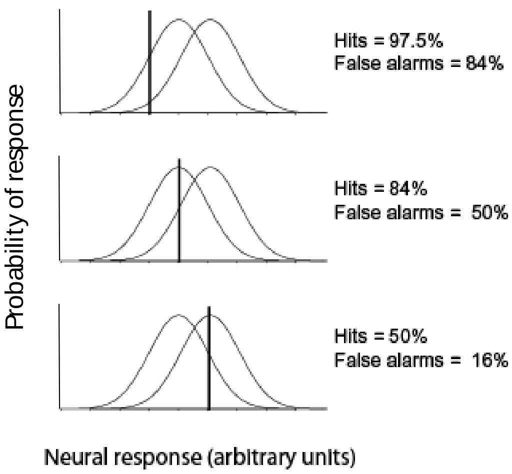


**Figure 3.4.** The effect of criterion changes on rates of hits and false alarms. In the top graphic, the observer has adopted a conservative response criterion (as indicated by the position of the vertical line) to maximise the proportion of hits. It can be seen that this has also resulted in a large proportion of false alarms. Note that even with such a low criterion for detection, the proportion of hits does not reach 100%; this is because of the Gaussian nature of the S and S+N distributions. From this position, a relatively small change in criterion (middle graphic) has the effect of a small reduction in the proportion of hits, but a much greater reduction in the number of false alarms. Finally, in the bottom graphic, the observer has raised their response criterion in order to minimise the proportion of false alarms. It can be seen that this has also caused a reduction in the proportion of hits – half of the stimuli are now undetected. Such a situation may arise where the cost of investigating a false alarm is prohibitive in comparison with the consequences of a miss. Adapted from http://www.cns.nyu.edu/%7edavid/handouts/sdt-advanced.pdf.

It can be seen from examination of the preceding section that signal detection theory describes how observers make decisions in the presence of uncertainty. The question arises, therefore: what type of decision/judgment does the observer need to make, and what does their response tell us? The examples so far have related to the detection of some sensory signal, a task which implies a response of either ‘yes, I detected the signal’ or ‘no, I did not detect the signal’. In fact, these response options are often problematic: a simple yes/no choice can be hugely influenced by observers’ response criterion. For example, the observer may adopt a very low criterion (in Figure 3.4, this would correspond to a vertical line representing criterion placed toward the left of the diagram) and frequently respond ‘yes’ when no stimulus is present (in this scenario, the proportion of false alarms will be high). This strategy may arise as part of a misplaced desire to ‘pass the test’ or perform ‘well’ in the experiment, and is even more likely if they expect an event to occur on every single trial. Clearly, if this is the case then the experimenter may erroneously conclude that the observer can see all of the stimuli, no matter how dim they are. This may lead the experimenter to run the experiment again with stimuli reduced in intensity, a wasteful course of action as the same confound will remain. One way to render such a criterion less likely is to have a proportion of trials without any stimulus presentation, where the observer should not respond. Such trials are known as ‘catch trials’. By emphasising to the observer before each run that accuracy is important, and that not all of the trials will contain a stimulus presentation, criterion can be raised to a more satisfactory level. This strategy can be even more effective if the observer knows that a false response conveys a penalty of some type (Rose, 2006).

### 3.1.2 Forced-choice paradigms

Given that criterion varies both between and within observers, it is highly desirable to use an experimental paradigm which minimises its effects on the measurement of the parameter of interest. A widely-used way of doing this is to use a ‘forced choice’ paradigm. Typically in such an experiment, an observer must choose between two possible responses, neither of which are ‘yes’ or ‘no’. Such a task is known as a two-alternative forced-choice (2AFC), two-interval forced-choice, or binary forced-choice paradigm; however, any number of alternatives can be used (e.g. 4AFC). The responses can be demarcated by space (spatial forced choice; e.g. the observer has to choose whether the stimulus appeared to the left or right of fixation) or time (temporal forced choice; e.g., the observer has to choose whether the stimulus appeared in the first or second temporal interval). The spatial location or temporal interval containing the stimulus (e.g., left or right, 1st or 2nd) is randomly determined on a trial-by-trial basis. Using such a forced-choice task, the effects of criterion variability are largely eliminated – an observer will not benefit from having a bias toward responding that (e.g.) the stimulus appeared to the left of fixation, as they will only be correct approximately 50% of the time. For this reason, forced-choice paradigms are generally considered to be a relatively criterion-free measure of perception.

Whether using a yes/no or forced choice task, at the end of data collection the experimenter must calculate the value of interest (e.g., detection threshold). By using a range of stimulus parameters spanning a perceptual continuum between ‘rarely detected’ and ‘easily detected’, observers’ responses can be plotted graphically and fitted with a curve – such a graph is often referred to as the ‘psychometric function’ of an observer for that task. Examples of typical psychometric functions for yes/no, 2AFC and 4AFC tasks are shown as Figure 3.5 (a-c).

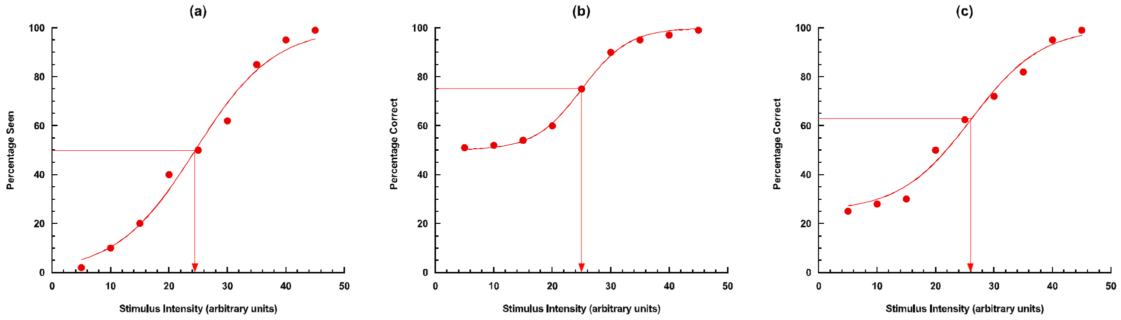


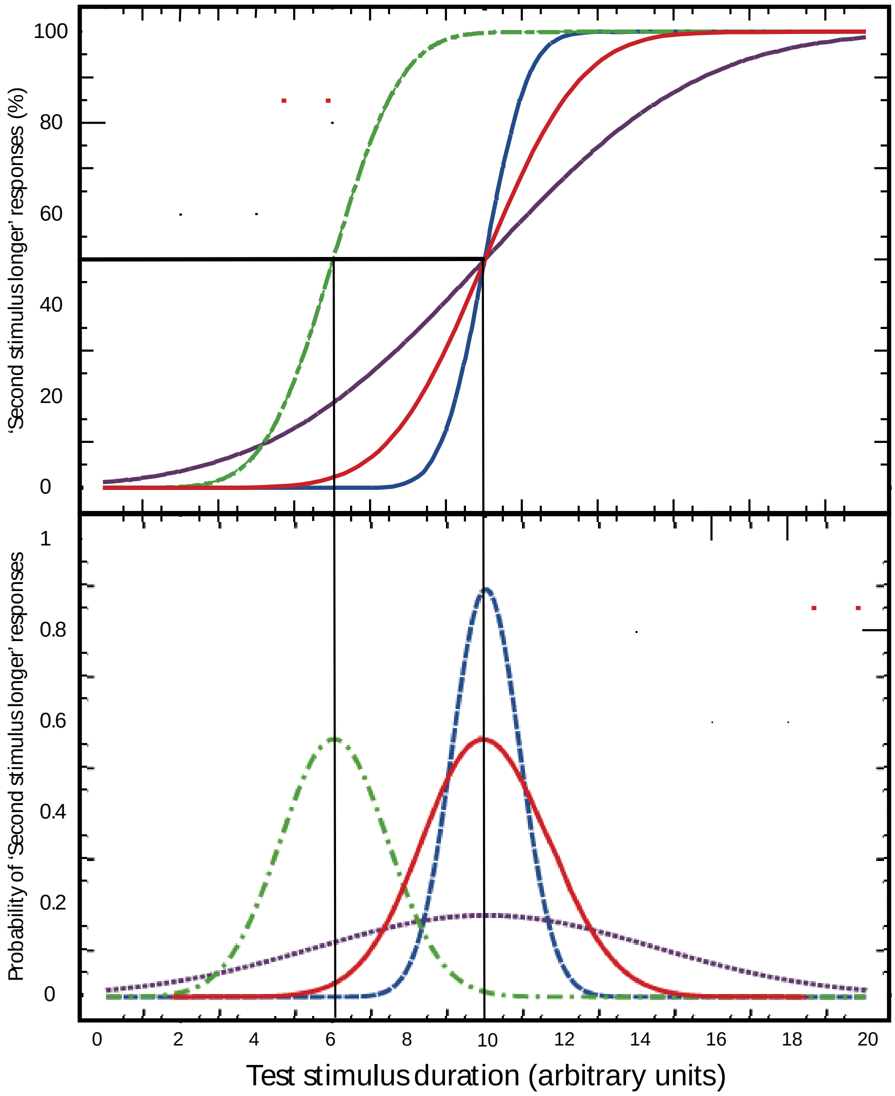
**Figure 3.5(a-c).** Figure 3.5(a) shows a standard psychometric function derived from the results of a visual detection task using a yes/no paradigm. Examination of the data points reveals that the lower the stimulus intensity, the lower the percentage of stimuli detected by the observer. The stimulus intensity corresponding to 50% detection is the estimate of threshold derived from the function. In Figure 3.5(b), detection threshold is measured with a 2AFC paradigm. As the observer has two response choices, performance would approximate 50% correct responses even if the observer did not see a single stimulus and guessed all the way through the experiment. In this case, the stimulus intensity corresponding to 75% correct responses is taken as the threshold estimate. Finally, Figure 3.5(c) depicts the psychometric function as derived from a 4AFC experiment. Here, the stimulus intensity corresponding to 62.5% correct responses (halfway between chance and perfect performance) is often taken as the threshold estimate.

The way in which the value of interest (e.g., stimulus intensity corresponding to detection threshold) is extrapolated from the psychometric function varies slightly between the different experimental paradigms. In a yes/no task, an observer simply has to state whether or not they detected the stimulus on any trial; the more intense the stimulus, the more likely the observer to discriminate it from the underlying noise (Figure 3.1). Conventionally, the stimulus intensity corresponding to detection on 50% of trials is taken as the measure of threshold (Figure 3.5 (a)), i.e. the observer is as likely to respond that they did see the stimulus as they did not see the stimulus. Thus, the simplistic notion of a fixed threshold level above which *all* stimuli are detected, and below which *all* stimuli are undetected, is unrealistic due to the omnipresent noise within the sensory system (Rose, 2006).

In a 2AFC task, however, 50% performance would not reflect an accurate measure of threshold as an observer could obtain this level of correct responses by chance or by simply pressing one button repeatedly and indiscriminately throughout the experimental run. It has already been seen in the preceding discussion that in a 2AFC task, an observer must respond as to which of two response dimensions (e.g., left or right) is correct. A more appropriate measure of threshold in this case would be the stimulus intensity corresponding to 75% correct responses (halfway between chance performance and 100% correct performance; Figure 3.5(b)). Similarly, in a 4AFC task threshold is usually taken as the intensity corresponding to 62.5% correct responses (halfway between chance performance of 25% correct and 100% correct responses (Figure 3.5(c)). The same principles can be applied to potentially any forced choice paradigm (e.g. 3AFC, 8AFC), although note that *any* value between chance and perfect performance may instead be chosen as representing threshold (Corliss & Norton, 2002).

It is important to note that although the preceding discussion has mostly used the example of a detection task measuring absolute threshold, the concepts and techniques outlined can be applied to *any* task measuring perception and/or sensation. Examples could include judging the relative spatial position of two vertical lines (to measure Vernier acuity), or judging the duration of one interval in comparison to another. In these cases, the 50% point on the psychometric function would not represent a measure of absolute threshold but the *Point of Subjective Equality* (PSE) – the point at which the test stimuli appear to be equal in the parameter of interest, e.g. spatial location or duration. A variant of PSE used extensively within this thesis is the *Point of Subjective Simultaneity* (PSS): the *physical* temporal offset between two stimuli required for *perceptual* simultaneity. This is discussed in more detail in Sections 3.3.1-3.3.3.

**Figure 3.6 (next page).** Illustration of the dependency of psychometric function slopes on observer sensitivity in a hypothetical duration discrimination task. In the top panel, the red, blue and purple functions all yield the same PSE value, where the two stimuli appear equal in duration (10 units), but are derived from observers with different sensitivity to changes in stimulus duration. The blue function is steepest, denoting the greatest sensitivity, and the purple is shallowest, denoting the least sensitivity. The slope of the red function denotes intermediate sensitivity, identical to the green function, but the two functions are horizontally displaced, denoting different mid-points and therefore different PSE values. The relationship between psychometric function slope and the underlying probability distribution is evident from inspection of the lower panel of the figure; the red and green probability distributions are of equal width but with peaks corresponding to different stimulus durations. Similarly, the blue and purple distributions have peaks corresponding to the same stimulus duration as the red distribution, but with different widths. Note that the area under the probability distribution functions is identical. Figure adapted from www.en. wikipedia.org/wiki/Normal_distribution.



Whereas the mid-point of the psychometric function gives the experimenter a measure of PSE, the slope of the psychometric function gives a measure of observers' performance at the task. A steep psychometric function is indicative of an observer who is relatively sensitive to a change in the stimulus parameters. In this case, a relatively small change in the stimulus parameter (e.g., an increase in the duration of a test stimulus) will cause a large change in the proportion of trials where the observer responds (e.g.) 'the test stimulus appeared longer than the standard'. Conversely, a shallow function indicates an observer with relatively little sensitivity to a change in the parameter of interest. Sensitivity/slope is intimately related to the width of the probability distribution function in signal detection theory (Figure 3.1), as the signal is easier to discriminate from the underlying noise if the probability distribution is narrow (Figure 3.2(a-b)). The slope of the function and its mid-point can change independently of one another. This is illustrated by Figure 3.6.

### 3.1.3 Curve fitting

The process of fitting a curve to data points is called a regression, and is a vital step in extracting the values of interest from the data. When performing the regression, it is important that all of the data points are incorporated into the psychometric function, including those which may not appear to convey much useful information (e.g., those where the observer is performing at close to 100% accuracy). In this thesis, all psychometric functions are fitted using the Method of Least Squares; although other methods exist, e.g. linear, quantile, and weighted least squares regressions, the method of least squares is a popular, relatively simple and well-understood method which is used by the software package KaleidaGraph (Synergy Software, U.S.A.), amongst others.

In the Method of Least Squares, the sum of the square of vertical distances between the function and each individual data point is minimised – the software computes a potentially infinite number of curve locations until it finds the one with the lowest total squared vertical distance from the data points. The distance between each data point and potential curve can be either positive (point above curve) or negative (point below curve), but the process of squaring these vertical offsets renders the polarity of the distances irrelevant. In the method of least squares, the curve does not need to pass through all of the data points; in fact, this situation is quite rare, and occurs only in cases of an exceptionally good fit of the curve to the data points. In this respect it differs from other methods of curve fitting, such as interpolation procedures, where the function intercepts each of the data points.

The method of least squares is appropriate for use with a small number of data points, which represents an advantage of the method when fitting data from psychophysical experiments. Because all of the data points are given equal weighting, however, a single outlying point can adversely affect the resultant regression. The chance of this occurring can be reduced by ensuring each data point represents a large number of presentations, and that all the points represent the same number of presentations. This is because although psychophysical measurements are inherently variable because of noise within the sensory system (Section 3.1.1), if the range of percepts associated with the stimuli is normally distributed then over the course of a large number of presentations outlying points should be reduced or eliminated.

In this thesis, all psychometric functions are fitted with a logistic function of the form

$$y = \frac{100}{1 + e^{-\frac{(x-\mu)}{\theta}}}$$

where μ is the mid-point of the resultant psychometric function and θ an estimate of the slope of the curve (related to, but not equal to, the function slope). These values need to be estimated by the experimenter prior to beginning the regression process, usually after inspection of the raw data (‘eyeballing’). The initial estimates form a starting point for the iterative calculations performed during the regression process. During this process the software package KaleidaGraph (Synergy Software, U.S.A), which is used throughout this thesis, computes a potentially infinite number of independent values for μ and θ using the Levenberg-Marquadt algorithm (Levenberg, 1944; Marquardt, 1963). The final values of these parameters are those which, when combined, minimise the least-squares variance of the data. Although other types of psychometric function exist, such as the cumulative Gaussian, probit, and Weibull functions, the logistic function is a relatively simple and widely-used method of curve fitting, and was judged suitable for use within this thesis. Using other types of function is likely to yield similar (although not identical) results (Strasburger, 2001).

In this thesis the majority of judgments made by observers concern the temporal properties of stimuli, more specifically the perceived temporal relationship between two stimuli. To maximise the usefulness of information obtained from observers’ responses, the range of stimuli presented and experimental methods employed need to be carefully chosen. Some of the classical experimental methods for measuring perceptual experience are summarised below.

## 3.2 Introduction to psychophysical theory and methods

Psychophysics may be defined as the study of the relationship between the physical properties of sensory events ('stimuli') and the resultant percept of an observer. By systematically varying the properties of the stimuli, and recording changes in observers' perception of the stimuli, the experimenter may infer the mechanisms by which such stimuli are processed within the central nervous system. Psychophysics is the oldest form of experimental psychology, and has been in use since the mid-19th century (Fechner, 1860). It is used extensively throughout this thesis. An overview of the most important methods is presented below.

### 3.2.1 The method of adjustment

The method of adjustment is arguably the simplest of psychophysical techniques, and can be used easily on observers without extensive periods of training. The observer has access to a control which decreases or increases stimulus intensity, and an experimental run typically commences with the intensity at either a grossly suprathreshold or subthreshold intensity (i.e., the observer is either able to perceive the stimulus with ease, or not at all). For example, if the experiment starts with a stimulus at a suprathreshold intensity level, the observer simply decreases the intensity until the stimulus *just* becomes imperceptible. If the experiment begins with a subthreshold stimulus, the observer must increase the intensity until it *just* becomes detectable. As the threshold obtained from ascending trials tends to be higher than that obtained when descending from a suprathreshold stimulus, typically both trial types are typically repeated many times and the results averaged to give an estimate of threshold.

This type of experiment has the advantage of being extremely quick and easy to perform, even for a naïve observer. However, the criterion used by an observer to decide whether or not they perceive a stimulus will vary enormously between individuals, with corresponding differences in results. Moreover, within observers, results also tend to be somewhat variable, most likely reflecting subjects changing their criterion throughout the course of the experiment. For these reasons, the method of adjustment is rarely used on its own to give an estimate of the parameter of interest. It may be useful, however, in giving an indication of the threshold region prior to testing using the method of constant stimuli (Section 3.2.4).

### 3.2.2 The method of limits

The method of limits is closely related to the method of adjustment; the major difference is that it is the experimenter or computer that adjusts stimulus characteristics between trials, rather than the observer. It is also typically used to measure detection thresholds. The experiment commences with stimulus intensity either clearly suprathreshold or clearly subthreshold, and decreases incrementally until the observer can no longer perceive the stimulus (descending series) or increases until the observer can just perceive the stimulus (ascending series). The average of a number of ascending and descending series is taken as an estimate of detection threshold.

It may be appreciated that this method can also be affected by observers' criterion in a similar fashion as the method of adjustment. For example, an observer who is eager to 'do well' in the experiment may adopt a lower criterion to report detection of the stimulus, and may respond that the stimulus is present when they are not yet certain. Additionally, observers may become accustomed to responding in a certain way during a long series of descending or ascending trials. For example, during a series of descending trials the observer may become used to responding 'yes, the stimulus was present' and continue responding in this way even when the stimulus is subthreshold (Corliss & Norton, 2002). Such a strategy will clearly give an erroneous estimate of threshold. Errors of this type are known as 'errors of habituation'.

If each ascending (or descending) experimental run begins at the same initial value of stimulus intensity, it is also possible that over the course of several runs observers may begin to anticipate the transition point based on their experiences of previous experimental runs. This may cause them to respond that they can perceive the stimulus before it is actually detectable (Corliss & Norton, 2002). Such an error is known as an 'error of expectation'. This causes bias in the estimate of threshold by a reduction of criterion. Such errors can be minimised by altering or randomising (within sensible limits) the starting point of the ascending/descending series between experimental runs.

Another potential criticism of the method of limits is that it is somewhat inefficient. The large proportion of trials that are grossly sub- or suprathreshold contribute very little useful information to the final estimate of threshold, as the observers' response on these trials is largely predictable; rather, they serve to increase the length of the experiment and contribute to observer fatigue. Only the 'transition' trials where the observer reports that the stimulus appears/disappears are used to calculate the parameter of interest. Clearly, a more efficient experimental paradigm would maximise the number of trials in the intensity region around the 'true' threshold, as observers' responses to these trials will give the experimenter the most useful information. Since the work of Fechner (1860), methods have been proposed to increase the efficiency and accuracy of the estimate of the parameter of interest by choosing stimulus parameters based on observers' responses up to that point in the experimental run. They are known variously as staircase procedures, the 'up and down method', and adaptive procedures. A discussion of these methods is presented below.

### 3.2.3 The staircase method

Consider the method of limits, as discussed above. At the transition point (e.g., where the observers' responses change from 'no, I did not detect the stimulus' to 'yes, I detected the stimulus'), the experimental run terminates, despite the fact that the stimulus is very close to the observers' detection threshold. In a staircase procedure, the run does not terminate at this point; rather, the ascending series of presentations turns into a *descending* series (or vice versa). This fundamental change in the nature of presentations (ascending to descending, descending to ascending) is referred to as a *reversal*. The descending presentations then continue until the next reversal (in this case, descending to ascending, i.e. until the observer reports that the stimulus is absent), and so on until a set number of reversals is reached and the procedure terminates. Before the beginning of each experimental run, the experimenter should have a rough idea of the expected threshold value and start the procedure at a similar level to ensure as few trials as possible are 'wasted' in a slow progression to threshold levels, such as in the method of limits. The procedure is illustrated in Figure 3.7.

The procedure summarised above is known as a standard staircase (Cornsweet, 1962). A disadvantage of this method is that, as with the method of limits, observers may make errors of habituation and expectation due to the predictable nature of step sizes and presentations (Cornsweet, 1962; Levitt, 1971). If the initial stimulus value is chosen poorly, it may also be inefficient due to the trials wasted in reaching the threshold region (Levitt, 1971).

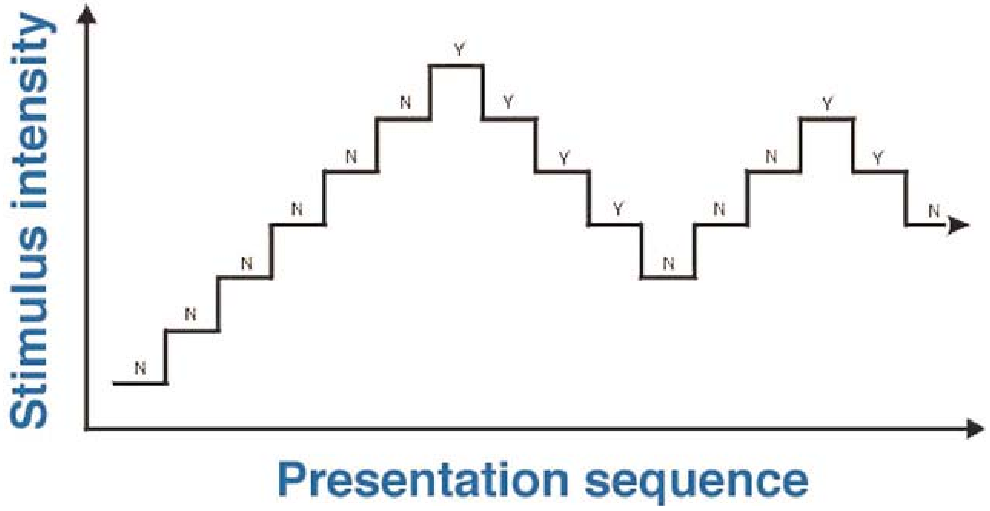


**Figure 3.7.** Illustration of the relationship between stimulus intensity and the observers responses (indicated by the small 'Y' or 'N' at each trial) during a hypothetical detection task using a standard staircase procedure. The procedure begins with a subthreshold presentation to which the observer responds that they did not perceive the stimulus. This response causes the stimulus intensity to increase, as in an ascending method of limits experiment. When the observer perceives the stimulus, intensity starts to decrease until it can no longer be seen again. This pattern continues until the run finishes. Taken from www.webvision.med.utah.edu/psych1.html.

However, many variations of the staircase method exist, such as the transformed staircase (Levitt, 1971), PEST (Taylor & Creelman, 1967), Best PEST (Pentland, 1980), QUEST (Watson & Pelli, 1983) and ML-PEST (Harvey, 1986). These employ varying rules or algorithms intended to increase efficiency and accuracy. The differences between the many various methods mostly relate to the size of the steps with which the parameter of interest (e.g., stimulus intensity) is changed between trials, when the changes occur, the termination point of the procedure, and the method for estimating threshold values at the end of the procedure.

For example, the original version of the PEST staircase procedure first described by Taylor and Creelman (1967) starts with a large step size which halves after each reversal of the staircase.  The point at which the stimulus level is changed (e.g., increases in intensity) is determined by a maximum-likelihood estimate of the value of interest (e.g., detection threshold) based upon *all* responses during the experimental run.  Therefore, in a PEST staircase the stimulus level does not automatically change on every trial – if the level is at or around the threshold, it is possible for a series of presentations to be made at the same level.  In other words, the procedure continuously tracks and updates an estimate of the value of the end-point throughout the experimental run.  Should the observer continue to either ascend or descend without reversal being reached, the fourth and subsequent steps in one direction double in step size to reach the reversal point using the minimum number of stimulus presentations.  These variations ensure that the procedure rapidly 'homes in' on the threshold region.  They also make it feasible for the staircase to begin with a random (within sensible limits) stimulus value to prevent anticipation of reversals by the observer, without significantly decreasing the efficiency of the procedure.

In the original version of the PEST algorithm, the procedure terminates when step size falls below a pre-determined value, with the final estimate of threshold being the last presented value.  The authors admit that this method of estimating the parameter of interest may be less precise than an averaging method (Taylor & Creelman, 1967); a subsequent variation on this sees the threshold estimate being derived from averaging the stimulus values (Kaplan, 1975). Such a procedure is known as a PEST (RAT mode). Other algorithms implement statistical estimation of the likely threshold based upon the responses throughout the procedure (Harvey, 1986; Pentland, 1980; Watson & Pelli, 1983). Due to the relatively complex nature of these estimation procedures, these methods tend to require more prior assumptions about the nature of the underlying psychometric function.

Despite the large number of variations of the staircase procedure, the relatively simple standard staircase is considered to give reliable results in an acceptably efficient manner, at least based upon results of computer simulations (Meese, 1995; Rose, Teller, & Rendlema, 1970) (although note that Pentland (1980) found that the Best PEST method gave less variable results). However, the possibility of randomising the initial stimulus values to prevent errors of habituation or expectation is undoubtedly a compelling reason to consider the use of a different algorithm. The huge variety of staircase procedures is only summarised here; the work of Treutwein provides a more detailed explanation of many more of the different algorithms and the estimation procedures used in them (Treutwein, 1995).

### 3.2.4 The Method of Constant Stimuli

The third of the psychophysical methods proposed by Fechner (1860) is the method of constant stimuli (MOCS). In this method, the stimulus presentations are made randomly across a fixed range of values. In a detection experiment, these values range from the grossly suprathreshold to the grossly subthreshold. The responses of observers reflect this varying difficulty, forming a continuum between presentations that are almost always detected to those that are almost never detected. Through presentation of a sufficiently large number of stimuli at each stimulus level, observers' proportion of responses (e.g., 'Yes, I heard the tone') can be plotted against stimulus strength to give a psychometric function (Section 3.1.2). The mid-point of this function (50% 'yes' responses) is the point of maximal uncertainty on the part of the observer, and represents the point at which they cannot decide whether they saw the stimulus or not. This level of intensity (or other variable) is taken as the measure of threshold.

Of course, rather than measuring detection threshold the experimenter may instead be interested in some other parameter, for example the physical duration of a stimulus that is perceived by to be equal in duration to another (different) stimulus (a duration discrimination task; Figure 3.6). To quantify this, an observer could be presented with two stimuli, one after the other. The first stimulus would be a 'standard' stimulus, with each presentation of the same duration. By presenting the second, 'test', stimulus at a range of durations and forcing the observer to choose whether the test stimulus is longer or shorter than the standard, a psychometric function can be plotted from the observers' responses. For example, when the test stimulus is grossly shorter than the standard, the observer will respond 'shorter' on the vast majority of trials, and vice versa. In this way, plotting the proportion of 'longer' responses against stimulus duration will give a psychometric function resembling one of those in Figure 3.6 (top panel). The test stimulus duration corresponding to 50% ‘longer’ responses is the PSE, the point at which the standard and test stimuli appear equal in duration.

As the order of presentations is completely random, observers are unable to speculate as to whether they are approaching threshold or not, and therefore errors of both habituation and expectation are eliminated. This represents a significant advantage over the method of limits. Similarly, the random order of presentations means that observers should regularly have to make relatively easy judgements, such as when the test stimulus is grossly longer/shorter than the standard. This may provide variety during the course of an experimental run, as well as reassuring the naïve or nervous observer that they are capable of performing the task. This contrasts with the later stages of a staircase procedure, where all presentations are made very close to the PSE; in this case, the judgements are all of a high level of difficulty, which may cause fatigue or boredom.

The main disadvantage of the method of constant stimuli is that a significant proportion of presentations are very easily discriminated or detected, e.g. a stimulus of duration significantly shorter than the standard, or a very bright light. Trials at these levels contribute very little to estimates of threshold (see also Section 3.2.2). In addition, many trials are needed in order to obtain reliable results. Preliminary experiments may also be necessary to obtain an estimate of the threshold so that a suitable range of stimulus intensities can be chosen. For these reasons, the method of constant stimuli is often regarded as somewhat inefficient, at least in relation to the adaptive methods (Watson & Fitzhugh, 1990). Despite this inefficiency, it is regarded as probably the most comprehensive and accurate method of quantifying the reactions of an observer to a stimulus (Corliss & Norton, 2002; Rose, 2006).

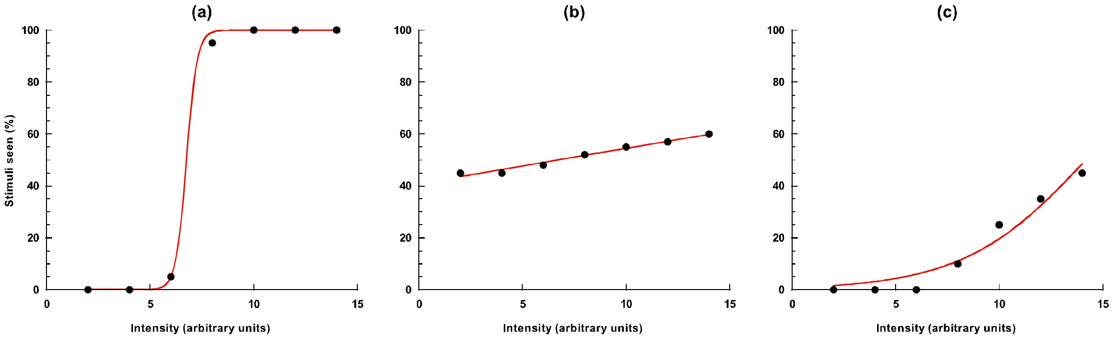


**Figure 3.8(a-c).** Illustration of the possible effects of unsuitable choice of stimulus range and values in a hypothetical single-interval visual detection experiment employing the method of constant stimuli. In Figure 3.8(a), the step size is too large for the observer. Rather than all the points contributing to the curve fit, the fit is influenced by only two of the seven points. The other points contribute little useful information to the graph, as they are either all detected or all not detected. The experiment should be repeated with smaller steps between the stimuli, such that the observers' responses form a perceptual continuum between 'almost always detected' and 'almost never detected'. In Figure 3.8(b), conversely, step size is too small, and the observer is unable to reliably detect any of the stimuli. This is seen by the way all the data points are clustered around the '50% stimuli seen' point. In Figure 3.8(c), the range of stimulus values chosen does not include the observers' likely threshold value and so 50% performance level is not reached. The experiment should be run again with the range of intensities tested shifted rightwards, centred on approximately 15 arbitrary units.

Key to data collection using the method of constant stimuli is the range of stimulus values and step sizes chosen for the experiment. Ideally, the point of interest (detection threshold, spatial offset for perceived alignment, duration of the test stimulus, etc.) should lie approximately in the middle of the stimulus range tested, in order to maximise the contribution of all the data points to the fit of the psychometric function. For the same reason, the step size chosen should be appropriate to observers' ability to perform the task. If these conditions are met, the range of values tested should span a perceptual continuum from (e.g., in a duration discrimination task) 'test duration almost always shorter' to 'test duration almost always longer' (Figure 3.6, top panel). The importance of stimulus values and step sizes is illustrated in Figure 3.8(a-c).

## 3.3 Methods of assessing relative timing perception in humans

Much of the data presented within this thesis pertains to observers' perception of the relative timing of two stimuli in different sensory modalities; vision, audition or touch. An introduction to some frequently used methods in assessing relative time perception is presented below.

### 3.3.1 Temporal order judgments (TOJs)

As the name suggests, in a temporal order judgment task an observer has to judge the temporal order of two stimuli. Observers are usually forced to make an unspeeded decision as to which of the stimuli came first. A range of time differences between the onset of the two stimuli, or Stimulus Onset Asynchronies (SOAs), are used, ideally spanning a perceptual continuum from one stimulus almost always being perceived first, to the other stimulus almost always being perceived first.  The order of SOA presentation is usually randomly determined via a method of constant stimuli.  The stimuli used can be of the same modality or of different sensory modalities; in this thesis, however, all temporal order judgments are made using stimuli in different modalities due to confounds frequently observed in within-modality TOJs (Section 2.1.3).

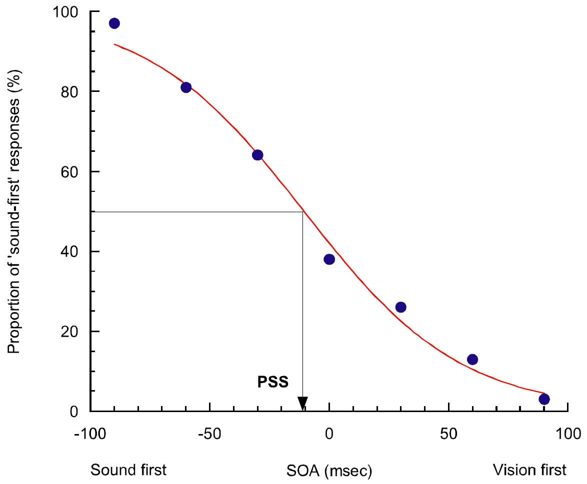


**Figure 3.9.** Sample psychometric function derived from the results of an audiovisual temporal order judgment (TOJ) task.  The stimuli are presented at a range of SOAs, with the observer forced to choose after each trial which of the two appeared to have come first.  It can be seen that with an increasing physical lead of sound over vision, the observer is more likely to respond 'sound first'.  Similarly, with an increasing lead of vision over sound, the observer is less likely to respond 'sound first'.  The physical temporal offset corresponding to 50% 'sound-first' responses is the PSS.

Following data collection, the proportion of (e.g.) 'sound-first' responses at each SOA are graphically plotted and fitted with a psychometric function. From this function, the SOA corresponding to chance (50%) performance is used as a measure of the Point of Subjective Simultaneity (PSS) – the *physical* temporal offset between the two stimuli corresponding to *perceptual* simultaneity.  The 50% point on the function is used because this denotes the physical temporal offset at which the observer is maximally uncertain as to which of the two stimuli came first (Figure 3.5(a) and Section 3.1.2). A sample psychometric function derived from an audiovisual temporal order judgment is shown as Figure 3.9.

In addition, the slope of the psychometric function can be used to quantify the sensitivity to temporal asynchrony of the observer, in the form of a Just-Noticeable Difference (JND). JND is simply a numerical estimate of the sensitivity of an observer to changes in the stimulus parameter of interest (Section 3.1.2), and is quantified throughout this thesis as half the offset between approximately the 27% and 73% points on the psychometric function (i.e., half of the SOA corresponding to ±1 $\theta$ from the PSS; Section 3.1.3). An observer with a high sensitivity to asynchrony will notice a small change in the physical temporal offset between the two stimuli and change their pattern of responses accordingly; hence, this observer will have a low JND and a relatively steeply sloping psychometric function (e.g., Figure 3.6, blue function). Conversely, an observer with a low sensitivity to asynchrony will need a large change in the temporal relationship before they significantly change their pattern of responses. Such an observer will have a high JND and a relatively shallow psychometric function (Figure 3.6, purple function).

### 3.3.2 Simultaneity judgments (SJ)

Simultaneity judgments differ from TOJs in that rather than judging the relative temporal order of the two stimuli, observers simply have to judge whether the stimuli are presented simultaneously or successively. As with TOJs, data collection involves the presentation of pairs of stimuli at varying SOAs, often using the method of constant stimuli. However, in simultaneity judgments the range of SOAs spans a perceptual continuum with ‘almost always perceived as asynchronous’ at opposite ends of the SOA range, with presentations in the middle of the SOA range more likely to be perceived as simultaneous. The proportion of ‘simultaneous’ responses are usually fitted with a Gaussian curve, with the peak of the Gaussian representing the SOA at which observers are most likely to perceive the two stimuli as simultaneous. The peak of the curve therefore corresponds to the PSS. A measure of the sensitivity of observers to asynchrony detection is provided by either the width of the Gaussian at half-height or its standard deviation, with a narrower function corresponding to a more sensitive observer. An example of a typical Gaussian derived from a SJ task is shown as Figure 3.10.

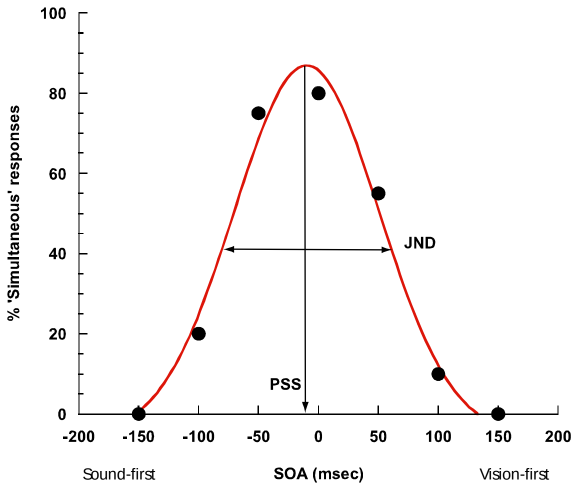


**Figure 3.10.** Gaussian curve fitted to hypothetical data derived from an audiovisual (AV) SJ task. The physical temporal offset between the two stimuli corresponding to the peak of the Gaussian is that at which the observer is most likely to respond that the two stimuli are simultaneous: the PSS. The width of the distribution of ‘simultaneous’ responses at half-height (or, alternatively, the standard deviation of the Gaussian) provides an estimate of observer sensitivity (JND).

### 3.3.3 Comparison of TOJ and SJ

As recently observed, in previous literature TOJs and SJs appear to have been used almost interchangeably as measures of perceived simultaneity (Vatakis, Navarra, Soto-Faraco, & Spence, 2008). Both of the tasks are potentially subject to observer bias and criterion, although in different ways. In a SJ, observers have to state whether or not two stimuli were presented successively or simultaneously. To make such a judgment, observers will have to adopt some criterion upon which to base their decision. As stated in Section 3.1.1, such criterion will naturally differ between observers, but can also vary within observers due to fatigue, mental state, and other factors. Consider the example of an observer with a low criterion for judging two stimuli as successive; such an observer will produce data with a large proportion of 'simultaneous' responses spread over a wide range of SOAs. Conversely, an observer with a high criterion for responding 'simultaneous' will produce data with a lower proportion of 'simultaneous' responses distributed over a narrower range of SOAs. Clearly, criterion will therefore affect both the *width* and the *amplitude* of the Gaussian response distribution. This confound will affect any estimate of observers' sensitivity to asynchrony, which is calculated from the width of the distribution at half-height. Importantly, however, criterion should not affect the *location* of the peak of the 'simultaneous' response distribution (Schneider & Bavelier, 2003). The estimate of PSS derived from a SJ task is therefore unlikely to be affected by observer bias and/or changing criteria.

In a TOJ task, conversely, it has been proposed that response bias *may* affect the location of the mid-point of the psychometric function, and therefore the PSS (Vatakis et al., 2008). For example, if a naïve observer is uncertain which of the two stimuli were presented first, they may adopt a strategy of responding that one of the stimuli always came first on trials where they cannot decide (i.e., when close to 50% performance levels). Such a strategy would shift PSS towards an SOA where that stimulus physically temporally lags the other stimulus. In such a case, the slope of the psychometric function (used to calculate the sensitivity of observers to changes to SOA) would not be altered. However, whether observers would indeed adopt such a strategy is untested. The proposed differing effects of observer bias and criterion on TOJ and SJ tasks has led one recent study to suggest that the appropriate task for use in a particular experiment may depend on the primary parameter of interest, PSS or observer sensitivity (Vatakis et al., 2008).

Despite these potential concerns, direct comparisons of TOJ and SJ data within the same study are few and relate only to the audiovisual (AV) sensory pairing. A recent study compared both PSS and observer sensitivity estimates obtained using both simple ‘flash’ and ‘click’ AV stimuli and a more complex visual display featuring a ball which appeared to ‘bounce’ against an on-screen object in temporal proximity to an auditory tone (van Eijk et al., 2008). The tasks employed were TOJ and SJ as described in Sections 3.3.1 and 3.3.2, respectively, and a combination of the two in which observers had to either respond that sound and vision were simultaneous, or choose which of the two appeared to come first (described by the authors as ‘SJ3’ due to the three response choices).  Results showed that TOJ PSS values showed no significant correlations with SJ or SJ3 values in both ‘flash/click’ and ‘bouncing ball’ displays, whereas the two SJ PSS estimates were highly correlated using both types of stimuli (van Eijk et al., 2008).  For each of the three tasks, PSS values were correlated between ‘flash/click’ and ‘bouncing ball’ displays. Measures of observer sensitivity to asynchrony also showed no correlation between TOJ and both types of SJ task using ‘flash/click’ stimuli, and a marginally significant correlation between TOJ and SJ (but not SJ3) sensitivity using the ‘bouncing ball’ display (van Eijk et al., 2008).

From their data, van Eijk and colleagues concluded that TOJ and SJ measure different aspects of observers’ perceived AV timing.  However, the lack of correlation between TOJ and SJ does not necessarily indicate which of the two measures provides the most truthful estimate of perceived relative timing. Additionally, the close correlation between PSS values measured with ‘flash/click’ and ‘bouncing ball’ stimuli applies to both TOJ and SJ (van Eijk et al., 2008); this suggests that estimates of PSS with both methods are robust and repeatable to some extent, irrespective of which one most accurately reflects relative sensory latencies.  In the future, it may be useful to undertake a more detailed comparison of the repeatability and robustness of TOJ and SJ PSS values over time and experimental manipulations.

It also appears that AV SJs may be more susceptible to the effects of adaptation to temporal asynchronies (discussed in Sections 5.1.1 and 5.3.4) than TOJs.  Vatakis and colleagues showed that exposure to film of human speech with the auditory soundtrack desynchronised by 300msec modulated AV PSS measured with a SJ, but not TOJ, task (Vatakis et al., 2008). Fujisaki and colleagues measured the effect of exposure to a series of asynchronous AV stimulus pairs on TOJs and SJs; although PSS was modulated in both cases, with TOJs "….the estimation was less stable and some participants did not show the adaptation effect" (Fujisaki et al., 2004). Whilst these results support the idea that TOJ and SJ may measure some different aspect(s) of relative timing perception, the question as to which most accurately reflects relative sensory latencies remains open, especially given the wide range of PSS values observed with both methodologies (Section 2.1.3). Overall, however, the argument that response criterion affects the estimate of sensitivity derived from SJ data could be considered more compelling than the argument that observer bias *may* modulate PSS in a TOJ task. Given the potentially important role of observer sensitivity in the proposed experiments, TOJs are used throughout this thesis.

### 3.3.4 The method of Libet

The timing method popularised by Libet and colleagues (Libet et al., 1983) has since been used by many authors to investigate the temporal dynamics of intention and/or movement, e.g. (Engbert & Wohlschlager, 2007; Haggard & Clark, 2003; Haggard et al., 2002; Lau, Rogers, Haggard, & Passingham, 2004; Lau, Rogers, & Passingham, 2006; Park et al., 2003; Wohlschlager, Engbert, & Haggard, 2003; Wohlschlager, Haggard, Gesierich, & Prinz, 2003). The paradigm, when employed, tends to vary remarkably little from the original experiment, so for the sake of brevity the description below is taken from the original work of Libet and colleagues (Libet et al., 1983). This study claimed to show that cerebral activity precedes an observers' reported intention to act by several hundreds of milliseconds, a finding which suggests that voluntary movements are (at least in part) initiated on an involuntary or unconscious basis (Libet et al., 1983).

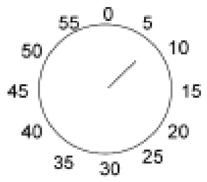


**Figure 3.11.** Illustration of a typical clock used in methods employing the method of Libet et al. (1983). The clock 'hand' (or alternatively a dot) rotates around the clock at a speed of one revolution every 2560msec. Following some predetermined event, the observer has to report the position of the hand/dot at event onset. Note that in the original experiment of Libet et al. (1983), the clock face had additional markings between each of the numbers (Haggard & Clark, 2003).

At the start of a typical experiment, the observer fixates the centre of a circle, the circumference of which contains markings corresponding to the 5-minute intervals on a standard clock face. Each '5-minute' interval is further divided into two, making a total of 24 divisions. A spot of light revolves around the edge of the screen (1.8° from fixation) at a rate of one revolution every 2560msec. The judgment made by the observer, whether it be the time when the observer first felt the urge to move (in the original experiments of Libet and colleagues), the time of a voluntary action, or the time of a sensory event, is recorded relative to the position of the spot on the clock. The physical position of the spot (as recorded by a computer) can then be compared with the position reported by the observer, and the difference between the two measurements can be computed to give a measure of any error in observers' perception of timing.  An illustration of the clock is shown as Figure 3.11.

Although the Libet method measures absolute time, it has frequently been employed to give a measure of the relative timing of two or more events.  This process may be exemplified by the work of Haggard and colleagues (Haggard et al., 2002).  In their study, Haggard et al. obtained measures of observers' percept of the timing of a voluntary keypress and a tone.  These measures were obtained in single-condition blocks, so that the tone could not be causally linked to the keypress.  In the 'operant' conditions, observers made a voluntary action (keypress) which was followed 250msec later by the tone; the task was to judge the timing of both these events in separate runs.  By calculating the difference between the perceived timing of events in the baseline and operant conditions, Haggard et al. (2002) were able to quantify shifts in the perceived timing of voluntary actions and subsequent auditory events; cause (keypress) and effect (tone) were shifted toward each other in perceived time.  These *intentional binding* effects are discussed in more detail in Section 2.1.6 and Chapter 7.

In spite of its widespread use, the paradigm has been the subject of criticism over the years, no doubt due (at least partly) to the extensive and ongoing debate relating to the conclusions of Libets' original paper (Libet et al., 1983). A substantial part of this criticism relates to the philosophical interpretation of Libet's results, which is unrelated to the present work and will not be discussed further; the interested reader may find the special issue of

Consciousness and Cognition, volume 11 (2) from 2002 a useful introduction to this ongoing debate.

Of more relevance to the present work as a potential source of error is the clock face used by observers as a reference against which to judge the position of the spot. As stated above, the spot revolves around the clock face at such a speed that each complete revolution takes 2560msec. As the clock face is marked with divisions corresponding to the five minute markers on a standard clock face (e.g. (Engbert & Wohlschlager, 2007; Haggard et al., 2002), each reference point used by the observer to make their judgment is 213msec apart. Therefore, the method imposes a minimum temporal resolution of 213msec upon observers; it is highly doubtful whether this will ensure a sufficient degree of precision in the judgment. Note that in the original experiment of Libet and colleagues (Libet et al., 1983) one additional point was placed between each of the '5 minute' markers, but this would still equate to 107msec between markers (Gomes, 1998). Other studies have instructed observers to give their judgements to the nearest 'minute' on the clock face (Engbert & Wohlschlager, 2007) or manually place the dot at the required position using a computer mouse or similar device (Lau et al., 2006), which in theory should improve the precision of observers' judgments.

As previously stated, observers using the Libet paradigm are required to judge some intention, action or event relative to a visual depiction of a clock face. As the visual input representing the clock face has to be transduced at the retina, processed in primary visual cortex (V1) and reach awareness before an observer can make any judgment relating to the clock (Section 1.1.1), it could be said that the *subjective* present (i.e., the position of the hand/dot on the clock face) is actually in the *objective* past (Hallett, 2007). If we assume a processing delay between a visual event occurring and becoming available to the perceptual system of up to 100msec (Klein, 2002), then any judgment made relative to the clock will have an error of up to 100msec associated with it.

When considering the implications of this error, it should be borne in mind that it may or may not affect results obtained with the Libet method in the visual modality; it has been speculated that this processing delay *may* be cancelled out or reduced by the flash-lag effect or some other factor(s), resulting in a veridical timing estimate from the observer (Klein, 2002; Pockett & Miller, 2007). The flash-lag effect refers to a phenomenon where a flashed stimulus presented physically aligned with a moving object appears to lag behind the moving object (for a review, see (Nijhawan, 2002)).

However, the latency of auditory and tactile events differs from that of visual events (Sections 1.1.1-1.1.3). Comparing the perceived time of a visual event measured using the Libet paradigm with that of (e.g.) an auditory event is therefore likely to be invalid, as different sensory latencies are likely to affect results in the various modalities in different ways. A recent study showed that this is in fact the case: using the Libet method, the error associated with estimating the event time of a tactile event differed significantly from that associated with time estimation of a visual or auditory event (Danquah, Farrell, & O'Boyle, 2008). This means that estimates of sensory event timing using the Libet method are subject to different errors depending on the modality of the sensory event, making comparisons of perceived event time between the modalities problematic. As this thesis is concerned with timing across the sensory modalities, such a confound is clearly highly undesirable.

Another potential criticism of the Libet paradigm centres on the role of attention. The Law of Prior Entry (Titchener, 1908) states that an attended stimulus is processed more rapidly than a similar unattended stimulus, and both behavioural (Shore et al., 2001; Spence et al., 2001b; Zampini et al., 2005c) and electrophysiological (Vibell et al., 2007) evidence exists to support this hypothesis. Should this be the case, it would have the effect of the attended object being processed earlier than normal, inducing an error in the timing judgement. This is important as in the Libet paradigm, an observer has to attend not only to their intention or action, but also to the timing apparatus itself, a clock face containing a revolving dot or hand. As the observer has to therefore attend to multiple perceptual streams, it is possible or even likely that attention may shift between the two aspects of the task in an uncontrolled fashion, rendering the results noisy and/or inaccurate.

Despite the potential criticisms of the Libet paradigm discussed above, a recent study concluded that the method yields accurate data regarding the perceived time of an observers' action, but only if a number of conditions are met (Pockett & Miller, 2007). However, due to concerns regarding the different timing errors associated with events in different sensory modalities

(Danquah et al., 2008), the method of Libet (Libet et al., 1983) was deemed unsuitable for use in this thesis.

# Chapter 4

## 4.1 An Investigation into the Relationship Between Reaction Time and Temporal Order

### 4.1.1 Introduction

Two of the principal methods used to investigate perceived timing in humans are the reaction time (RT) and temporal order judgment (TOJ) tasks. RT measures the time taken for an observer to detect and respond to a sensory event (Section 2.1.1). A TOJ task, conversely, requires observers to judge which of two sensory events appeared to come first; the *physical* temporal relationship between the stimuli required for *perceptual* simultaneity is a measure of the Point of Subjective Simultaneity (PSS) (Sections 2.1.3 and 3.3.1).

Given that RT and TOJ both represent measures of sensory latency, we may expect that RTs in different sensory modalities may predict TOJ results for those modalities. For example, the RT literature suggests that RT to a visual stimulus is slower than that to an auditory stimulus by 40-50msec (Section 2.1.1). It therefore seems reasonable to assume that in an audiovisual (AV) TOJ task, a visual stimulus must physically lead an auditory stimulus for perceived simultaneity. However, as reviewed in detail in Section 2.1.3, the evidence to support such a hypothesis is mixed: although many studies have reported that a physical lead of vision over sound is indeed required for perceived simultaneity, many other studies have also reported that a physical lead of sound over vision is necessary for perceived simultaneity. The available literature in the visuotactile (VT) domain is similarly dichotomous in its findings. Conversely, in the audiotactile (AT) pairing, the currently available literature appears unanimous in finding that a physical lead of touch over sound is necessary for perceived simultaneity. This corresponds to tactile RT being slower than auditory (Section 2.1.1). The reason(s) for these seemingly contradictory findings remain unclear.

Resolution of this issue from examination of the available literature is difficult as the majority of studies measured TOJs alone, with only two studies (both in the AV domain) also measuring RT with the same experimental apparatus and parameters (Jaskowski et al., 1990; Rutschmann & Link, 1964). This is relevant because if sensory latency (as measure by RT) is able to predict perceived temporal order, then PSS should equal the algebraic difference between simple RTs for any given modality pairing for each observer. Although the relationships between RT in different modalities are relatively well-established, a significant degree of inter-study and inter-observer variability exists. Thus, only by directly comparing PSS measured by TOJ with that predicted by RT for each observer is it possible to conclude whether or not PSS can be predicted by differences in RT. Measuring RT and TOJ using the same apparatus and parameters is important because it is well-established that RT can be modulated by factors such as stimulus intensity and duration (Cattell, 1886; Diederich & Colonius, 2004; Froeberg, 1907;

Hildreth, 1973; Jaskowski & Sobieralska, 2004; Kammer et al., 1999; Kohfeld, 1971; Pins & Bonnet, 2000; Ulrich et al., 1998). Although PSS can also be affected by these variables, the effect on RT is larger (Jaskowski, 1992; Roufs, 1974). Therefore, any comparisons of RT and PSS values obtained from different studies using different apparatus, stimulus parameters and observers are likely to be invalid.

With these factors in mind, the present series of experiments had the following aims:

- To measure simple RTs in the visual (V), auditory (A) and tactile (T) modalities for the purposes of comparison to subsequent measurements.

- To measure crossmodal PSS for AV, AT and VT stimulus pairings, using a TOJ task and the same apparatus and stimuli used in the RT measurements to facilitate direct comparison with the RT data. Examination of the results should be able to answer whether a dissociation between RT and TOJ exists in the AV and/or VT pairings, as suggested by previous work (Section 2.1.3), and whether previous results suggesting no such dissociation in the AT pairing can be replicated.

## 4.2 Measurement of Simple Reaction Time

### 4.2.1 Subjects

Subjects for this experiment were author JVMH, DW, JH and naïve observer DS. All had normal or corrected-to-normal vision, no reported auditory or tactile sensory impairments, and all had extensive previous experience of RT measurement.

### 4.2.2 Methods and Stimuli

The visual stimulus consisted of a 10msec flash of a 10mm diameter green Light Emitting Diode (LED) positioned at a viewing distance of 550mm (equivalent angular subtense of 1.05°), which was powered by the parallel port of a Dell desktop PC (www.dell.co.uk). The LED was held at approximately eye level by a metal stand, and produced a visual stimulus with a luminous intensity of 600cd/m$^2$.

The auditory stimulus consisted of a 10msec square-wave windowed white noise burst delivered via the sound card of the PC binaurally through Sennheiser HD650 linear headphones at 70dB SPL (measured with Bruel & Kjaer type 2250 sound level meter). The relative timing of the electrical input to the headphones and the auditory output of the headphones (measured via a piezoelectric actuator placed on the speaker cone) was recorded on a dual storage oscilloscope. This confirmed that the headphones produced the auditory stimulus approximately simultaneously (delay <1msec) with the electrical input, and eliminated the headphones as a potential timing error source.

The tactile stimulus was provided by a 24V dc push-pull solenoid (www.rswww.com), which delivered a ‘tap’ to the forefinger of the left hand. As the solenoid was driven by the serial port of the PC, which would ordinarily deliver insufficient current for successful operation, a Manson EP-907 power supply (Manson Engineering Industrial Ltd., Hong Kong) was used in conjunction with a custom-designed circuit to meet the power requirements of the stimulus. The solenoid tip took 5msec to reach the forefinger, and this small delay was subtracted from all tactile reaction times. The solenoid was driven by a 15msec signal from the serial port of the PC which, allowing for the 5msec rise time, provided a ‘tap’ of 10msec duration as confirmed by storage on an oscilloscope[2].

Although the stimuli were all clearly suprathreshold, no specific attempt to match the intensities of the stimuli was made. As remarked in a previous study, it is not currently clear which criterion is the most appropriate to match stimuli in different modalities: subjective intensity, detection latency, discrimination latency, or some other criterion (Spence et al., 2001b). Therefore, although there is abundant evidence to suggest that RT in the

---

[2] During verification, the action of the solenoid caused an electrical circuit to be completed and the resulting electrical current produced the signal for storage on the oscilloscope.

visual (Cattell, 1886; Kammer et al., 1999), auditory (Kohfeld, 1971; Ulrich et al., 1998) and tactile (Diederich & Colonius, 2004) modalities can be modulated by stimulus intensity, the present series of experiments used unmatched, suprathreshold stimuli, as in the majority of other studies (e.g. Boulter (1977)).

A problem inherent in the use of both solenoids and vibrotactile devices as tactile stimuli is that of noise generated during operation. This is important as if the operational noise of a tactile transducer is audible to observers, the resultant ‘tactile’ RT measurement may in fact resemble an auditory measurement. In previous studies, this noise has been masked by low-level background white noise played over headphones (Boulter, 1977; Spence et al., 2003; Spence et al., 2001a; Zampini et al., 2005a) or by physical damping of the apparatus (Diederich & Colonius, 2004). The present study employed physical damping by placing a small felt disc between the two metal contact surfaces of the solenoid, with a central hole in the felt to allow the passage of the central pin. The solenoid was mounted in a plastic cylinder, which was topped with 8mm of rubber to absorb vibration when the two surfaces made contact; again, a central hole allowed the free movement of the central pin and enabled it to make contact with the index finger. Finally, the subject placed their arm inside a cylinder fashioned of multiple layers of heavy fabric, inside which was also placed the tactile stimulus (on a sheet of foam rubber). The combination of this physical damping and the pinna-enclosing headphones ensured that subjects were unable to perceive any operational noise from the solenoid.

Measurement with the oscilloscope revealed that the time lag between pressing a mouse button and the response being registered by the PC was approximately 25msec[3]. If uncorrected, this delay would cause all recorded RT values to be 25msec slower than observers' physical responses to the stimuli; therefore, this systematic error was compensated for within the computer programme controlling the experiment. All stimuli were controlled by custom-written software run in MatLab (Mathworks, U.S.A.) on the Dell PC described above.

### 4.2.3 Procedures

The subject sat in a darkened room in front of the PC, with their right hand on a computer mouse[4]. Each modality was tested in separate blocks. When testing the visual modality the subject fixated the unlit LED before commencing an experimental block. Headphones were worn before the start of a run testing the auditory modality, whilst prior to testing the tactile modality the subjects rested their left forefinger on the solenoid, their whole forearm being enclosed in the previously-described heavy fabric cylinder, whilst wearing the headphones.

---

[3] Measurement was accomplished by using a piezoelectric actuator placed beneath the mousebutton; when the button was pressed, a current was generated which was stored on the oscilloscope for simultaneous comparison with the stimulus onset time.

[4] One observer (JH) was left-handed, and it remains feasible that responding with his unfavoured (right) hand caused RT to be slightly slowed. However, absolute RT values were

Each experimental run consisted of 45 stimulus presentations in the relevant modality, with an inter-stimulus interval (ISI) which varied randomly (with a uniform probability) between 250 and 750msec. Subjects were instructed to attend to the modality in question, and press the computer mouse button as quickly as possible when they perceived a stimulus. The first five presentations of each run were used as practice and therefore excluded from data analysis. As in previous RT studies (Jeeves & Moes, 1996; Romei, Murray, Merabet, & Thut, 2007), RTs shorter than 100msec were treated as anticipatory in nature; observers were advised of this fact and instructed not to respond until they were certain that an event had occurred. On the rare occasions that a sub-100msec trial was registered, the experimental run was terminated and observers repeated the run. Each observer completed five experimental runs in each condition (visual, auditory, and tactile), making a total of 600 test presentations (40 presentations * 5 runs * 3 modalities). The condition changed at the end of each experimental run in order to prevent any possible practice effects affecting the modalities differentially.

The data were collated and analysed using KaleidaGraph v3.5 (Synergy Software, U.S.A.).

---

of secondary interest to *relative* values, which would have been equally affected in the different modalities in the event of RT being slowed.

## 4.2.4 Results

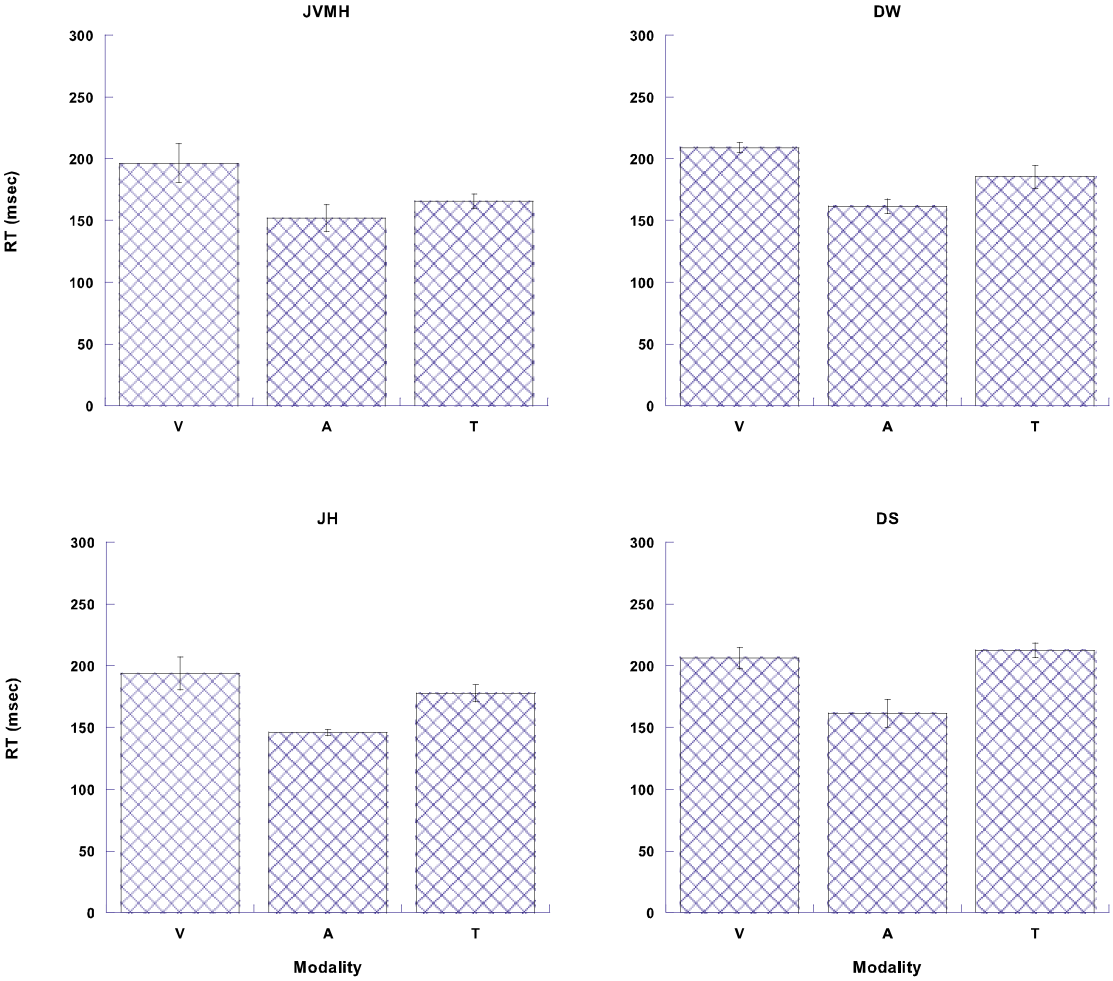


**Figure 4.1.** Graphs showing mean simple visual (V), auditory (A) and tactile (T) RT in milliseconds for all four observers. Error bars represent one standard deviation (SD) either side of the parameter values; the standard error of the mean (SEM) can be calculated from the equation SEM = SD/√N, where N is equal to the number of measurements from which the mean and SD are derived.

### 4.2.5 Discussion

Visual, auditory and tactile RTs for all observers are comparable to the majority of previous studies, as summarised in Section 2.1.1. The difference between A and V RT, approximately 45msec for each observer, is also comparable to the majority of the literature. From inspection of the RT data for each observer, it becomes possible to predict the results of TOJs for any given stimulus pairing. For example, from the V and A data obtained from author JVMH, one would predict that the visual stimulus would have to be presented 45msec before the auditory stimulus for the two to be perceived as simultaneous. To test this hypothesis, a crossmodal TOJ task was devised.

## 4.3 Measurement of crossmodal PSS using a TOJ task

### 4.3.1 Subjects

Participants in this experiment were the same as those from Section 4.2.1. As before, subject DS was naïve as to the purpose of the experiment. All observers had extensive previous experience of a TOJ task.

### 4.3.2 Methods and Stimuli

Visual, auditory and tactile stimuli were the same as those used to measure simple RTs, as described above. The extremely brief (10msec) nature of the stimuli means that it remains possible that observers used the relative stimulus offset, as well as onset, timing to aid their TOJ in spite of the instruction to judge the onset order of the stimuli. Previous work eliminated this potential confound by keeping the auditory and visual stimuli switched on until the observer made their SJ or RT response (Stone et al., 2001); however, this would result in all stimuli being of different durations, and was considered undesirable given the effects of stimulus duration on RT and PSS (Section 4.1.1).

The stimuli were controlled by custom-written software run in MatLab (Mathworks, U.S.A.). Each pair of stimuli was presented at one of seven physical temporal offsets, or Stimulus Onset Asynchronies (SOAs): -90, -60, -30, 0 (simultaneous), 30, 60 and 90msec, which were randomly interleaved within a method of constant stimuli (Section 3.2.4). Positive and negative asynchronies are used to denote which stimulus was presented first, positive always referring to a visual lead and negative always to a tactile lead (for a more detailed explanation, see Section 4.3.4). The relative timings of all stimuli were verified by simultaneous capture on a dual storage oscilloscope.

### 4.3.3 Procedures

The subject sat in the same darkened room as used for the RT measurements. The three stimulus pairings (AV, AT and VT) were tested sequentially (e.g., one run of AV, followed by one of AT, one of VT, and so on), to reduce or eliminate the effect of practise on the planned analysis of JNDs. An experimental run consisted of 75 presentations, the first five being practice presentations that were excluded from data analysis, and ten presentations at each of the seven SOAs. Each subject completed five experimental runs, making a total of 350 valid presentations (50 at each SOA) per modality pairing. The subject was instructed to attend to both modalities to be tested, and make an unspeeded judgment as to “which modality occurred first”. The subject responded by clicking a mouse with their right hand for one of the stimuli (this key was always used to signal a ‘touch first’ response), or pressing a key on a computer keyboard for the other (this key was always used to signal a ‘vision first’ response).

Results were collated and analysed using KaleidaGraph v3.5 (Synergy Software, U.S.A.). The percentage of sound-first (AV trials) or touch-first (AT and VT trials) responses was plotted as a function of stimulus onset asynchrony (SOA) and fitted with a logistic function of the form

$$y = \frac{100}{1 + e^{-\frac{(x-\mu)}{\theta}}}$$

where μ is the level of asynchrony equal to the PSS ( corresponding to 50% auditory-first or tactile-first responses on the psychometric function) and θ provides an estimate of asynchrony detection threshold or 'Just-Noticeable Difference' (JND; Section 3.1.2) (approximately half the offset between the 27% and 73% response levels on the psychometric function).

### 4.3.4 Results

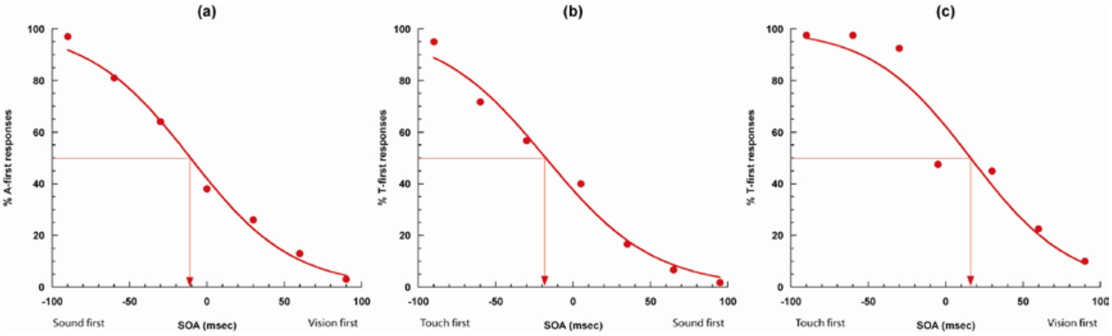


**Figure 4.2(a-c).** Psychometric functions derived from crossmodal TOJs for observer JVMH in the AV, AT and VT pairings (a-c, respectively). In each case, the percentage of 'sound-first' (a) or 'touch-first' (b-c) responses are plotted as a function of the physical temporal offset (SOA) between the stimuli. Negative values correspond to either an auditory lead (AV condition) or a tactile lead (AT and VT conditions). Inspection of the functions shows that observer JVMH requires sound to be presented approximately 11msec before vision for perceived simultaneity (the SOA corresponding to chance performance), touch to be presented 17msec before sound, and vision presented 16msec before touch. An implicit measure of the fit of the function to the data points is given by the error bars representing standard deviation in Figures 4.3-4.5.

Psychometric functions derived from crossmodal TOJs for observer JVMH are shown in Figure 4.2(a-c, corresponding to the three modality pairings). The physical temporal offset corresponding to the mid-point of the functions (indicated by the arrow) is the PSS. For example, inspection of Figure 4.2(a) shows that author JVMH requires an auditory stimulus to be presented approximately 11msec before a visual stimulus for the two to be perceived as simultaneous.

The RT results from Section 4.2.4 can be used to predict PSS values in each modality pairing, with predicted PSS being the algebraic difference between RT values. In this way, it is possible to plot predicted PSS against measured PSS for each observer. This information is shown in Figures 4.3-4.5.

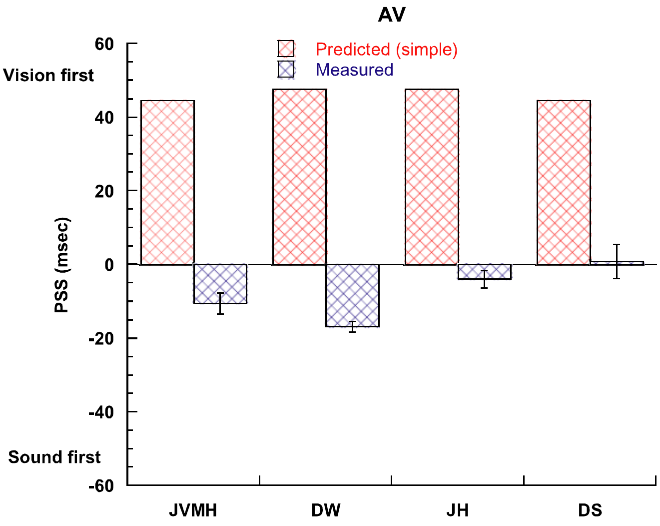


**Figure 4.3.** Comparison of AV PSS values predicted from simple RT values (red bars) with those measured using a TOJ task (blue bars) for all four observers. Error bars represent one standard deviation either side of the parameter values[5].

From the data in Figure 4.3, it can be seen that for observers JVMH and DW, an auditory stimulus needs to be presented slightly *before* a visual stimulus for the two to be perceived as simultaneous in a TOJ task (note that observer JH also requires an auditory lead for perceived simultaneity, but that this PSS

---

[5] Error bars are not shown for predicted PSS (red bars) as the data were subjected to an ANOVA (described in detail in the main text), and thus an estimate of predicted RT variability from combined RT SDs could be considered superfluous.

does not differ significantly from zero (p > 0.05))[6]. This is the opposite result as would be predicted from simple RT as measured in Section 4.2.4. For observers JH and DS, perception of AV temporal order is approximately veridical, i.e. physically simultaneous auditory and visual signals are perceived as simultaneous. This also represents a dissociation between RT and TOJ, as PSS ≠ RT(V)-RT(A). In summary, simple RT cannot be used to predict PSS for all four observers in the AV pairing.

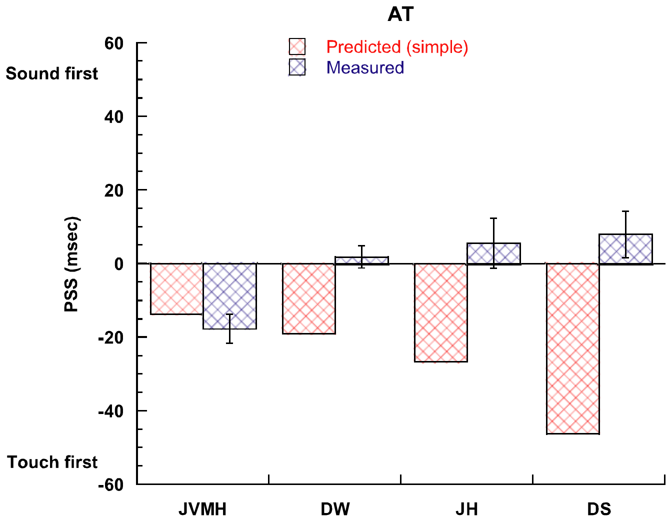


**Figure 4.4.** Comparison of AT PSS values predicted from simple RT values (red bars) to those measured using a TOJ task (blue bars) for all four observers. Error bars represent one standard deviation either side of the parameter values.

Upon examination of Figure 4.4 it is apparent that observers DW, JH and DS all exhibit an approximately veridical percept of AT temporal order. This is despite the fact that RTs to auditory stimuli are faster than those to tactile stimuli (Sections 2.1.1 and 4.2.4). Therefore, these observers exhibit a dissociation between RT and TOJ similar to that found in the AV pairing. In contrast, for author JVMH PSS is not significantly different from RT(A)-RT(T) (p > 0.05).

---

[6] This is evidenced by the fact that measured PSS is within 1.96 SD of zero.

Figure 4.5 indicates that all four observers demonstrate a dissociation between RT and TOJ in the VT pairing; measured PSS does not equal the difference between RTs. This difference is significant for all observers ($p < 0.05$). All observers require a visual stimulus to be presented before a tactile stimulus for the two to be perceived as simultaneous, although for JH this visual lead is statistically non-significant ($p > 0.05$). In summary, RT is unable to predict PSS for all four observers.

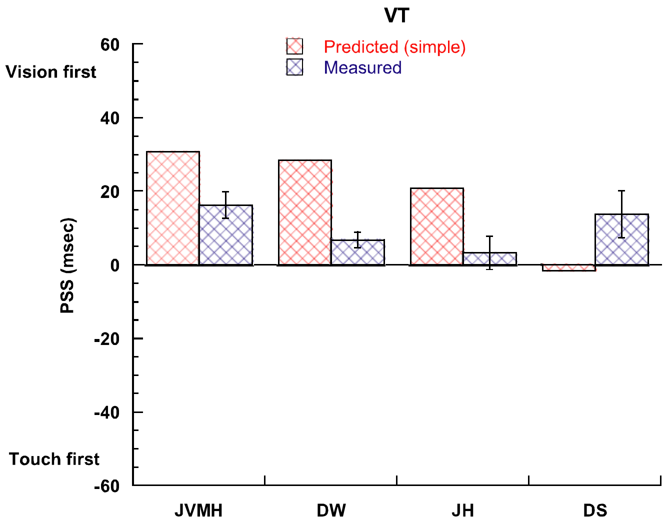


**Figure 4.5.** Comparison of VT PSS values predicted from simple RT values (red bars) to those measured using a TOJ task (blue bars) for all four observers. Error bars represent one standard deviation either side of the parameter values.

From the preceding discussion, it can be seen that clear discrepancies are evident between PSS as predicted by RT and as measured with a TOJ task. This can be visualised in the form of a scatter plot. Figure 4.6 plots predicted PSS against actual PSS. If simple RT can be used to accurately predict perceived temporal order, then the data points should be positioned on or around the diagonal (gradient = 1). Note that only one of 12 points occupies such a position. If, conversely, the nervous system compensates for differential neural processing latencies across modalities to maintain a veridical perception of synchrony, then the data points should be positioned around the horizontal (where measured PSS = zero). Six points out of 12 meet such criteria at the 95% confidence level. The overall picture is somewhat equivocal, and the only conclusion that can be drawn from this data is that simple RT cannot be used to predict perceived temporal order as measured with a TOJ task. This was confirmed by analysing the correlation between predicted and measured PSS in the AV, AT and VT pairings; the results of this analysis are shown in Table 4.1.

| Sensory pairing | Linear correlation coefficient | p |
|---|---|---|
| AV | -0.38 | 0.62 |
| AT | -0.76 | 0.25 |
| VT | -0.18 | 0.82 |

**Table 4.1.** Correlation and p-values for PSS predicted by simple RT and measured using a TOJ task.

Returning to Figures 4.3-4.5, it can be seen that the magnitude of the observed RT/TOJ dissociation (that is, the amount by which algebraic RT difference (predicted PSS) differs from measured PSS) appears to differ between modality pairings; most noticeably, the dissociation appears greatest in the AV pairing and smallest in the VT pairing. A repeated-measures ANOVA confirmed that the differences between the three pairings was statistically significant ($F_{2,11} = 7.09$, $p < 0.05$); however, subsequent Tukeys HSD post-hoc analysis showed that the only combination of pairings which significantly differed were the AV and VT pairings ($p < 0.05$), with the other two comparisons failing to reach significance ($p > 0.05$).

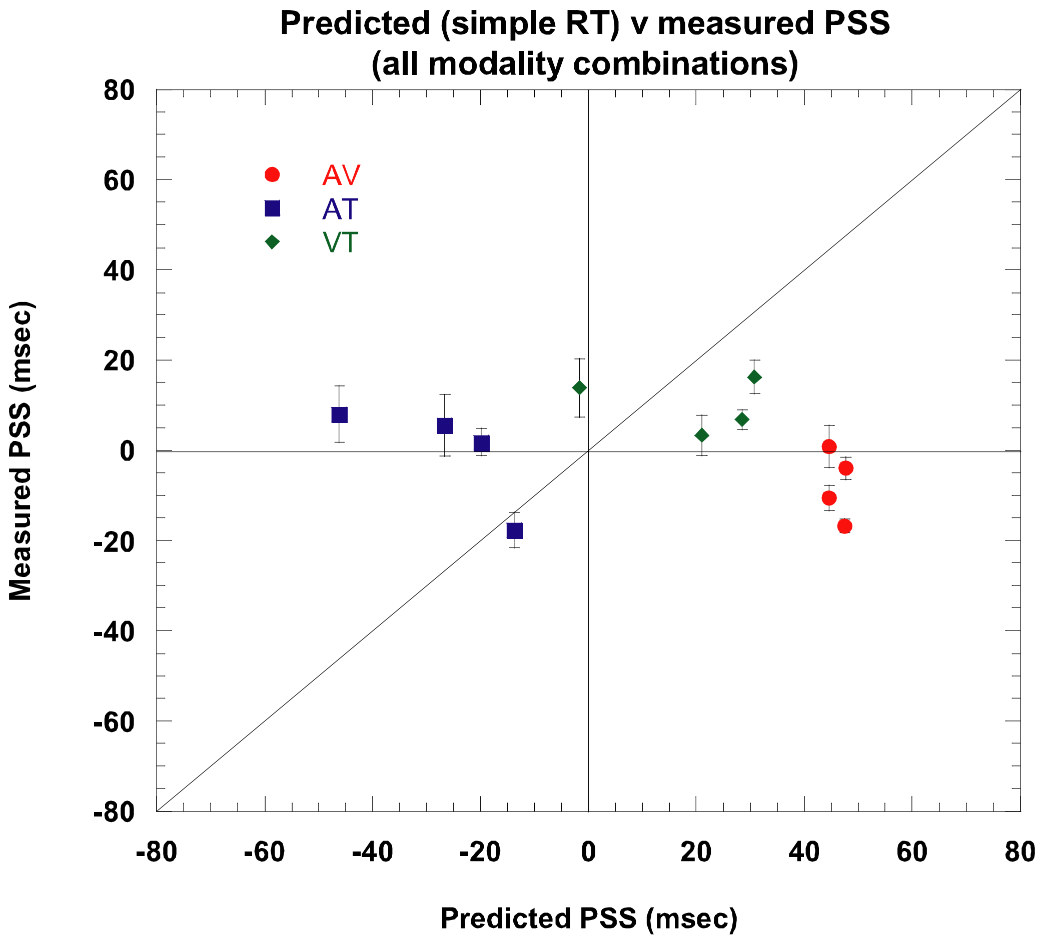


**Figure 4.6.** Scatter plot showing predicted vs. measured PSS for all modality combinations and all four observers. The X-axis shows predicted PSS, as given by algebraic RT differences for each modality pairing (measured in Experiment 1). The Y-axis shows PSS measured with a TOJ task. AV values are shown by red circles, AT by blue squares, and VT by green diamonds; individual data points represent individual observers. If simple RT is able to accurately predict PSS, values should lie on or around the diagonal line (gradient = 1), where predicted PSS is equal to measured PSS. If the perceptual system is able to compensate for intermodal differences in latency and RT, then data points should lie on or around the central horizontal line, where measured PSS = 0. Error bars represent one standard deviation either side of the parameter values.

Figure 4.7(a-c) shows JND for each observer in the AV, AT and VT pairings respectively. Examination of the figures reveals varying sensitivities between observers, with observer DW appearing the most sensitive to asynchrony in all modality pairings; this is most likely to be because this observer has the most experience of psychophysical procedures of the four subjects. A repeated-measures ANOVA indicated that within-observer JNDs did not differ significantly between the three modality pairings ($F_{2, 6} = 2.25$, $p > 0.05$).

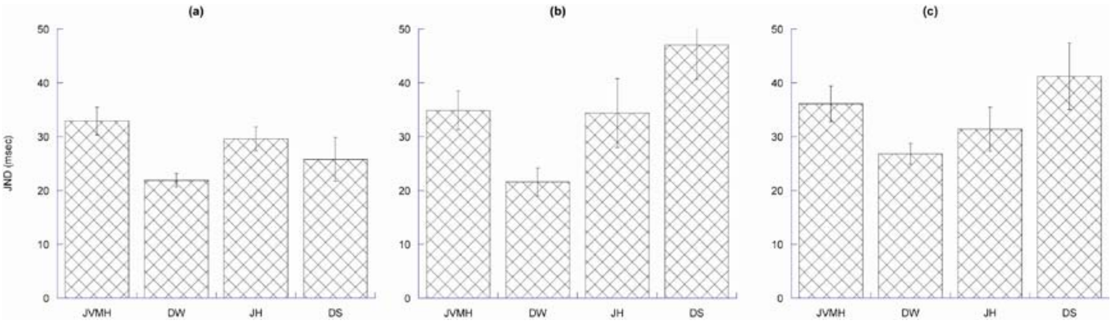


**Figure 4.7(a-c).** JND values for each observer in the AV, AT and VT pairings (a-c, respectively). Error bars represent one standard deviation either side of the parameter values.

### 4.3.5 Discussion

Results show that for observers JVMH and DW, an auditory stimulus needs to be presented before a visual stimulus for the two to be perceived as simultaneous. This result is in agreement with other studies (Arnold et al., 2005; Engel & Dougherty, 1971; Fujisaki et al., 2004; Hamlin, 1895; Harrar & Harris, 2005, 2008; Heron et al., 2007; Navarra et al., 2005; Neumann et al., 1992; Rutschmann & Link, 1964; Smith, 1933; Sugita & Suzuki, 2003; Teatini et al., 1976), but in opposition to others which indicate a visual lead is required (Dinnerstein & Zlotogura, 1968; Exner, 1875; Hirsh & Sherrick, 1961; Jaskowski et al., 1990; Keetels & Vroomen, 2005; Kopinska & Harris, 2004; Lewald & Guski, 2004; Machulla et al., 2007; Spence et al., 2003; van Eijk et al., 2008; Vatakis et al., 2007; Zampini et al., 2003a). The necessity of an auditory lead for perceived simultaneity directly contradicts the reaction times presented in Section 4.2.4.

Given that stimulus intensity and saliency can modulate both RT (Cardoso-Leite, Gorea, & Mamassian, 2007; Kohfeld, 1971) and (to a lesser extent) *unimodal* PSS (Cardoso-Leite et al., 2007; Jaskowski, 1992; Roufs, 1974; Sanford, 1974), it seems possible that the unmatched intensities of the stimuli used in the present study (and others) may have contributed to the pattern of results. However, a closer examination of the literature reveals that other studies using stimuli matched for intensity have also obtained opposing results; Rutschmann & Link (1964) and Neumann et al. (1992) found that an auditory lead is required for perceived simultaneity, and Jaskowski et al. (1990) found that a visual lead is required. For observers JH and DS, the PSS does not differ significantly from zero, indicating that simultaneous presentation of audition and vision is perceived as approximately simultaneous. This result also has precedent in the TOJ literature (Heron et al., 2004). It should be noted, however, that observers JH and DS still display an RT/TOJ dissociation as $PSS \neq RT(V) - RT(A)$. The discrepancy between TOJ results for observers JVMH and DW, and JH and DS, is compatible with other studies into perceived AV simultaneity (e.g., Stone et al. (2001)) which have suggested that the PSS may be observer-specific. This suggestion is also consistent with the wide range of AV PSS values obtained in different studies, recently summarised by van Eijk et al. (2008).

In the AT pairing, a similar dissociation between RT and TOJ is present for observers DW, JH and DS. Previous work (Dinnerstein & Zlotogura, 1968; Harrar & Harris, 2008; Hirsh & Sherrick, 1961; Machulla et al., 2007; Navarra et al., 2007; Zampini et al., 2005a) indicates that a tactile lead is required for perceived simultaneity in this modality pairing; however, in the present study this was only true for author JVMH, the other three observers requiring a small (statistically non-significant) auditory lead. Therefore, the results of the present experiment are mostly in opposition to those obtained elsewhere. However, as RTs were not also measured in these other studies, it is difficult to draw any firm conclusions about the relationship between RT and AT TOJ from this body of work.

In the VT pairing, all observers show a dissociation between RT and TOJ: PSS ≠ RT(V) –RT(T). Quantitatively, the magnitude of the dissociation between RT and TOJ is lower in this pairing than the other two pairings tested. Qualitatively, the VT results retain a certain degree of compatibility with previous work which also concluded that a physical lead of vision over touch is necessary for perceptual simultaneity (Dinnerstein & Zlotogura, 1968; Machulla et al., 2007; Poliakoff et al., 2006; Spence et al., 2003; Spence et al., 2001b). However, the results of the current experiment go beyond this and clearly show that simple RT cannot predict perceived temporal order in this modality pairing. The reason(s) for the discrepancy with other TOJ studies which claim that a physical tactile lead is necessary for perceived simultaneity (Harrar & Harris, 2005, 2008; Hirsh & Sherrick, 1961; Occelli et al., 2008b) is unclear.

Examination of JNDs obtained in the present study reveal them to be broadly similar to those recorded in other studies, e.g. ((Adelstein, Begault, Anderson, & Wenzel, 2003; Spence et al., 2003; Zampini et al., 2005a), although a detailed comparison is problematic because of differences in experimental methods and the amount of training undergone by observers. However, the fact that quantitative analysis reveals JND to be approximately the same in the AV, AT and VT pairings (despite clear differences in the temporal processing capabilities of the individual modalities (Section 2.1.2)) may suggest that a single, supramodal neural mechanism mediates crossmodal TOJs. As discussed in Section 2.1.3, the neural locus of such a mechanism is currently unknown. The only previous study to compare JND in the three modality pairings (Hirsh & Sherrick, 1961) was unable to present any statistical analysis to support their observation that psychometric functions in the AV, AT and VT conditions appeared to have similar slopes to each other as well as to unimodal V, A and T functions (indicating similar JND values). However, the conclusion of this previous study is clearly in agreement with the present work.

As stated previously, why such a dissociation between simple RT and TOJ should exist is unclear at present. It is evident, however, that although RT and TOJ tasks both represent measures of sensory latency, attentional requirements differ between the two tasks: in a simple RT task, the subject need only attend to the single modality being tested, whilst in a TOJ task, the subject has to attend to two modalities simultaneously. The effect of attention on sensory latency is well-documented – i.e., the ‘prior entry’ phenomenon discussed in Section 2.1.3. Given that latency may be modified by attention, the question arises as to whether division of attention between the sensory modalities may affect latency values such that differential RT values correspond to PSS as measured with a TOJ task. Such a situation could arise if observers were unable to divide their attention equally between the sensory modalities when instructed to do so.

In support of this hypothesis, previous work has claimed that sensory input from the visual modality can ‘dominate’ that from other modalities, i.e. when instructed to divide attention equally between modalities, visual performance and awareness appears to be affected less than that of other modalities (Colavita, 1974; Egeth & Sager, 1977; Hartcher-O'Brien et al., 2008; Hecht & Reiner, 2009; Klein, 1977; Koppen & Spence, 2007a, 2007b, 2007c, 2007d; Posner, Nissen, & Klein, 1976; Sinnett, Spence, & Soto-Faraco, 2007; Ward, 1994; Ward, McDonald, & Lin, 2000). These effects are often referred to as *visual dominance*, and have been interpreted as an unconscious bias of attentional resources towards visual sensory input (Posner et al., 1976). Given that sensory latency is frequently found to be modulated by attention (Section 2.1.3), it would be unsurprising if RTs measured when attention is divided between sensory modalities differ from simple (unimodal) RT values. Any such differences may enable RT measured under conditions of divided attention to predict PSS measured using a TOJ task, and explain the dissociation of RT and TOJ. If attention were to be divided unequally (whether consciously or unconsciously) between the modalities then sensory latency would most likely be affected in a similarly unequal fashion. Previously, inter-observer differences in perceived simultaneity have been suggested to reflect similar differences in the distribution of attentional resources between modalities (Sanford, 1888b).

With these factors in mind, RT was measured again with attention divided between sensory modalities. One of two possible outcomes was anticipated. Firstly, relative latencies (as measured by RT) under conditions of divided attention would change to reflect PSS values; for example, for author JVMH we would expect that auditory RT would be longer than visual by approximately 11msec. Such an outcome would show that RT and TOJ are equivalent measures of perceptual latency when attention is divided between modalities in a similar manner. Alternatively, it is possible that RT measured under conditions of divided attention would be unable to predict PSS as measured with a TOJ task. Such an outcome would suggest that the two tasks measure some fundamentally different aspect(s) of sensory latency.

## 4.4 Measurement of Reaction Times under Conditions of Divided Attention

### 4.4.1 Subjects

Subjects were the same as in the previous experiments (Sections 4.2.1 and 4.3.1). As before, observer DS was naïve as to the purpose of the study.

### 4.4.2 Methods and Stimuli

Visual, auditory and tactile stimuli were the same as used in the previous two experiments. All stimuli were controlled by custom-written software run in MatLab (Mathworks, U.S.A.), which also compensated for the 25msec delay in registering observers' responses caused by the computer hardware and operating system. As well as testing the three possible bimodal stimulus combinations used in the TOJ task, a trimodal condition using all three stimuli was included. This allowed quantification of the effects of three levels of stimulus certainty (100%, 50% and 33%) on RT.

### 4.4.3 Procedures

The subject sat in the same darkened room used for the previous experiments. Each bi- and trimodal experimental run consisted of 85 and 125 presentations respectively (40 per modality), with the first five presentations being used as practise and not included in data analysis. Each modality was presented an equal number of times within an experimental run, the precise order being generated at random. It must be emphasised that although the experimental blocks are described as bi- or trimodal, this only refers to the number of possible modalities per block; each presentation was made in one modality only. The ISI varied randomly (with a uniform probability) between 250 and 750msec. The subject was instructed to distribute their attention equally between the modalities to be tested, and respond as soon as they perceived *any* stimulus by pressing a computer mouse button with the right hand. As with the unimodal trials (Section 4.2.3), RTs shorter than 100msec caused the termination of the experimental run. After an initial practice run (data not analysed) in each condition, each subject completed five experimental runs in each condition (AV, AT, VT, trimodal), making a total of 1800 valid presentations per observer (40 presentations * 5 runs * 3 conditions per modality * 3 modalities). As with the unimodal conditions, the different modality pairings were tested sequentially.

The data were analysed in three categories for each modality: ‘mean’ RT (average of all presentations for that modality), ‘same’ RT (average of all trials where the preceding trial was in the *same* modality) and ‘diff’ (different) RT (average of all trials where the preceding trial was in a *different* modality). This was done in order to ensure that the Modality Shifting Effect (MSE) would not affect the pattern of results; the MSE quantifies the cost to RT performance of shifting attention between the sensory modalities. Specifically, the presentation of a stimulus in one modality is thought to exogenously direct attention toward that modality so that the processing of a subsequent stimulus in a *different* modality is delayed (Turatto, Benso, Galfano, & Umilta, 2002). Such attentional shifts have previously been shown to affect both response latency in a choice RT task (Spence et al., 2001a) and response accuracy (Hseih, 2002). In the present experiment, a MSE would be apparent if ‘same’ RT values were significantly faster than ‘diff’ RT values.

### 4.4.4 Results

Figures 4.8 - 4.11 summarise the data for all four observers. From inspection of the Figures, it appears that visual RT is slightly slower under conditions of divided attention than when attention is focussed only on the visual modality. A one-way repeated-measures ANOVA confirmed that a significant difference between the visual conditions was present ($F_{3,15} = 7.84$, $p < 0.01$); however, Tukeys HSD post-hoc comparison revealed that this significance arose due to a difference between ‘alone’ and trimodal conditions ($p < 0.005$), with no other differences reaching significance ($p > 0.05$).

For all observers, RT to auditory stimuli appears to be significantly slower under conditions of divided attention than when attention is focussed on the auditory modality. A one-way repeated measures ANOVA showed a highly significant difference between the conditions ($F_{3,15} = 17.78$, $p < 0.001$), and a Tukeys HSD post-hoc comparison revealed a significant difference between ‘alone’ and all other conditions ($p < 0.05$); modality uncertainty thus results in slower auditory RT.

For all four observers, no significant difference in tactile RT is evident in *any* of the modality combinations tested, including the simple RT measurements from Section 4.2.4 ($F_{3,15} = 0.529$, $p > 0.05$).

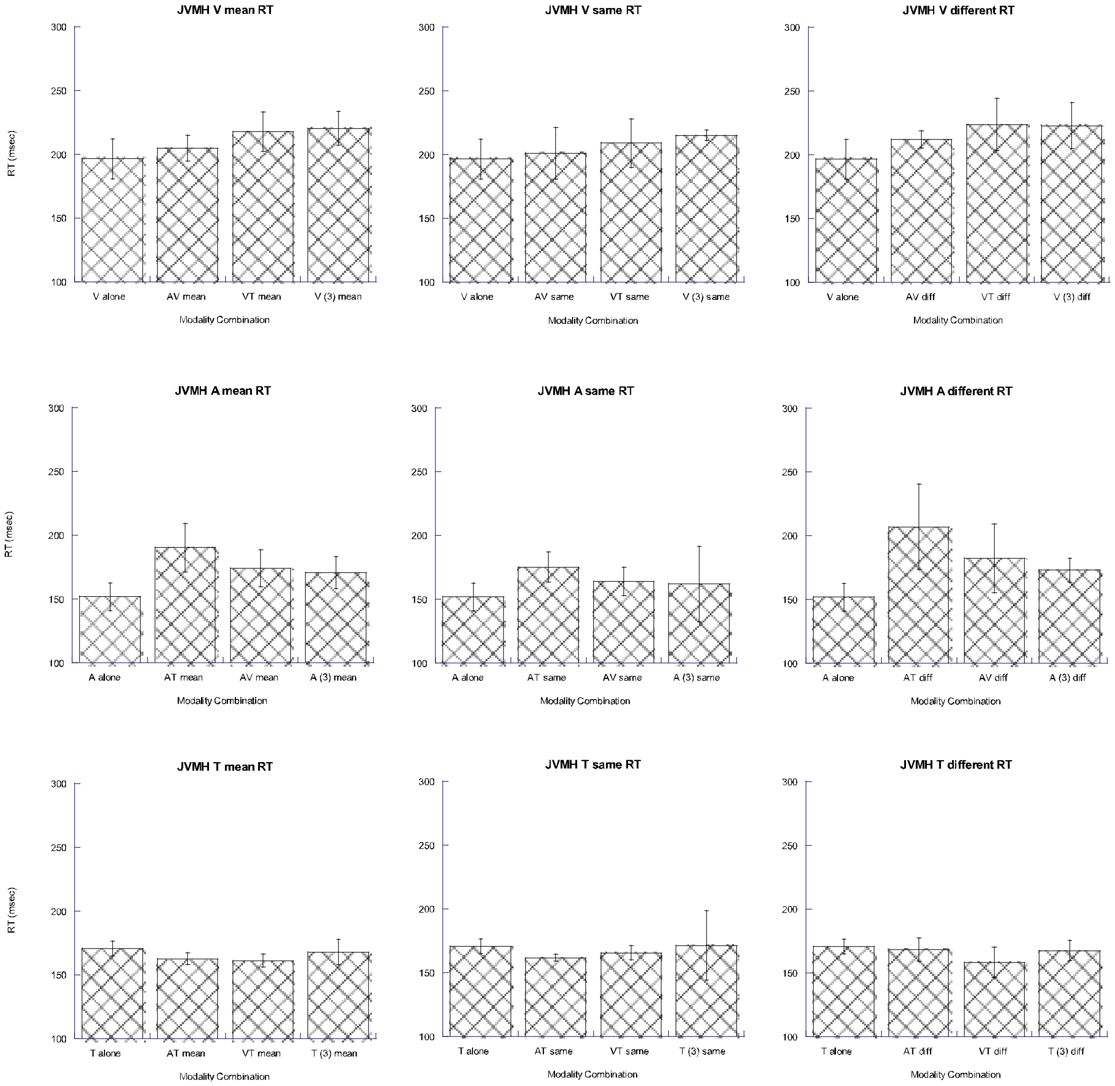


**Figure 4.8.** Plots showing RTs measured under conditions of divided attention for observer JVMH. The nine plots are arranged in three rows by modality (visual (top), auditory (middle), and tactile bottom)) and three columns by trial type (‘mean’ trials (left), ‘same’ trials (middle), and ‘diff’ trials (right). ‘Mean’ bars show the average of all trials for that modality in that condition, ‘same’ trials the average of trials where the preceding presentation was in the same modality, and ‘diff’ trials the average of trials where the preceding presentation was in a different modality. On each of the nine plots, the leftmost bar (‘alone’) shows simple RT for the relevant modality as presented in Section 4.2.4, and the rightmost bar (‘3’) shows RT measured whilst attention was divided between all three modalities (i.e., when modality certainty was 33%). The middle two bars show RT for the bimodal combinations of that stimulus (modality certainty of 50%). Error bars represent one standard deviation either side of the parameter values.

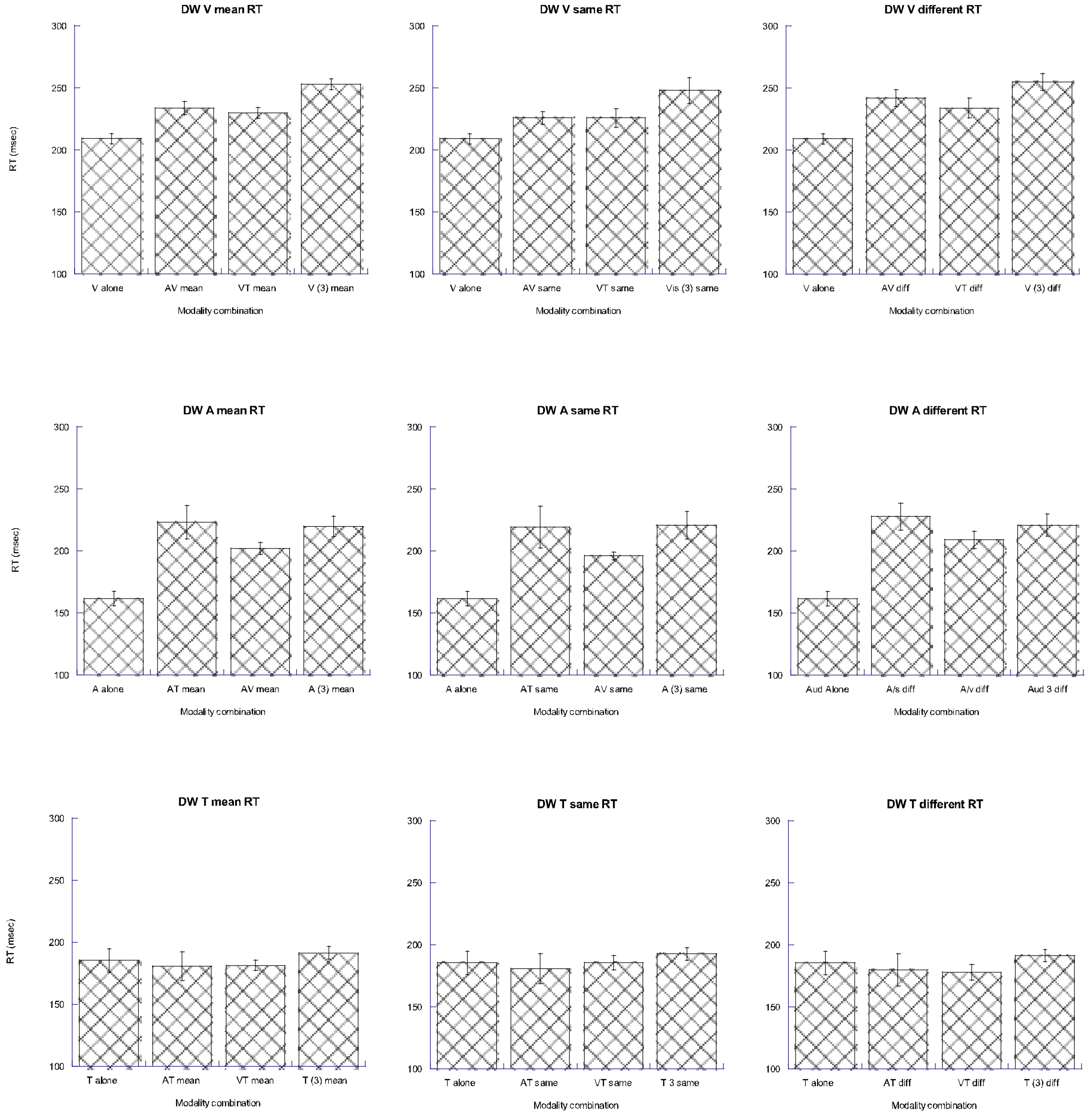


**Figure 4.9.** Figure showing RTs measured under conditions of divided attention for observer DW. The arrangement of graphs is identical to that in Figure 4.8. Error bars represent one standard deviation either side of the parameter value.

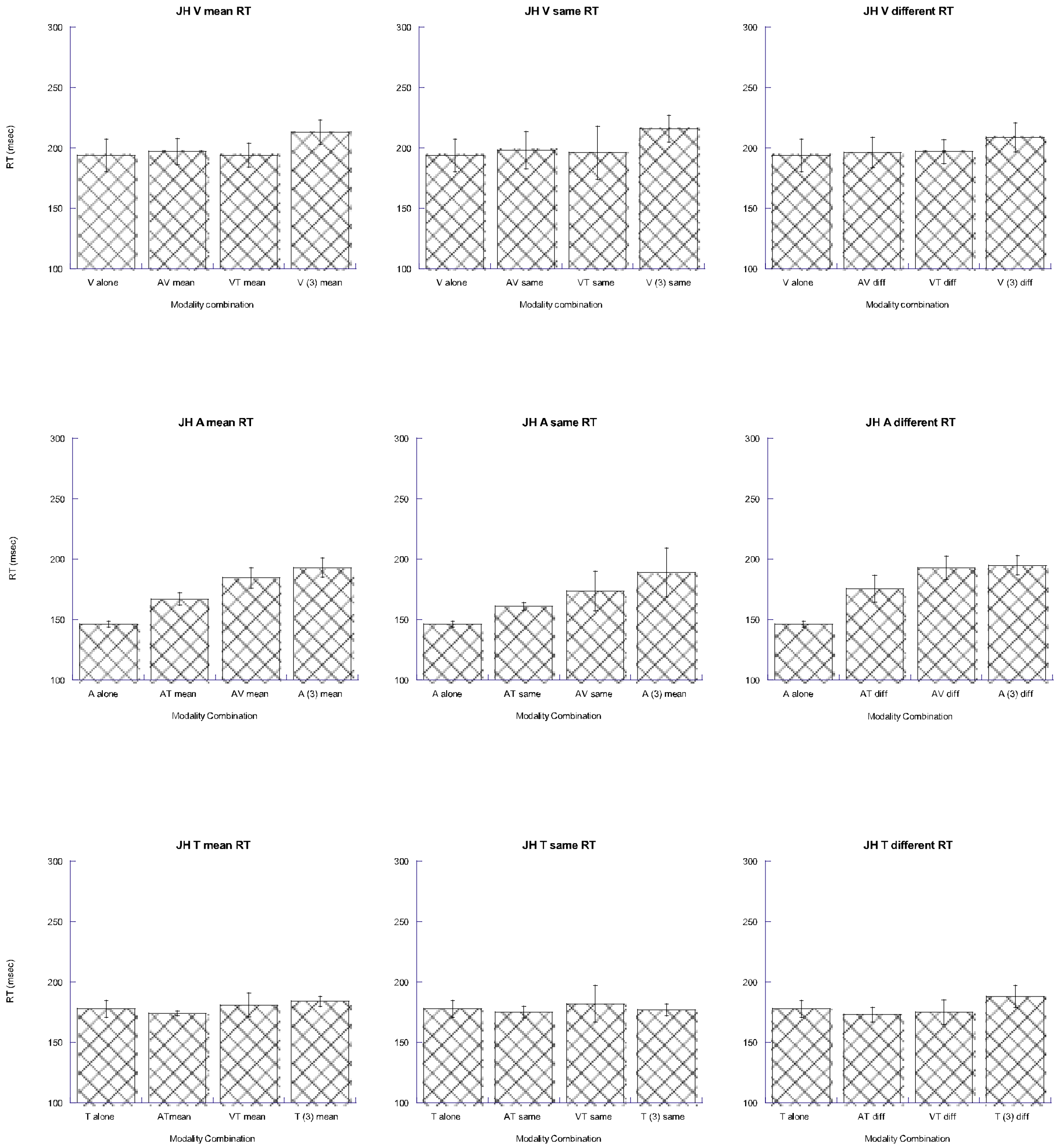


**Figure 4.10.** Figure showing RTs measured under conditions of divided attention for observer JH. The arrangement of graphs is identical to that in Figure 4.8. Error bars represent one standard deviation either side of the parameter value.

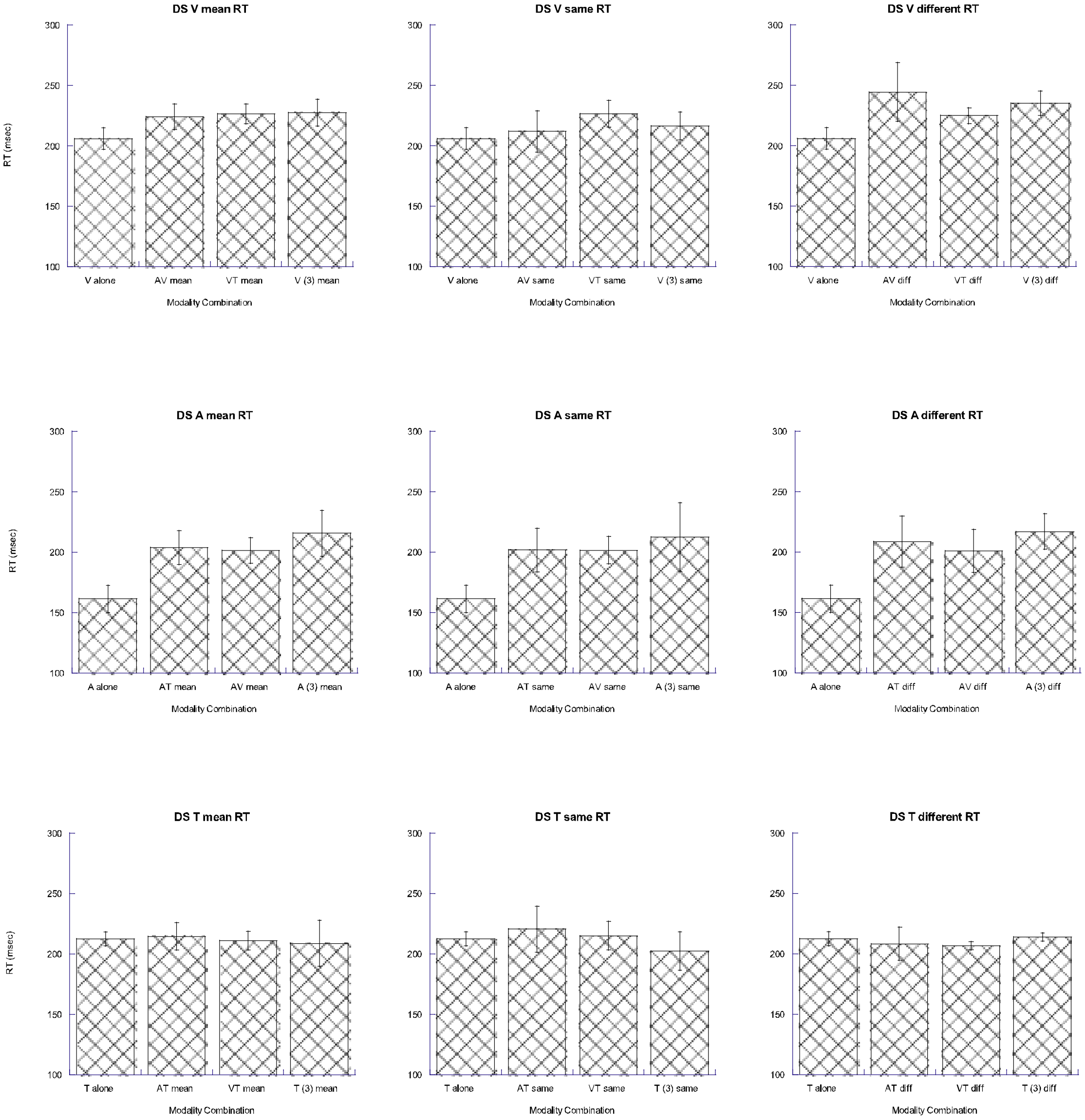


**Figure 4.11.** Figure showing RTs measured under conditions of divided attention for observer DS. The arrangement of graphs is identical to that in Figure 4.8. Error bars represent one standard deviation either side of the parameter value.

Examination of Figure 4.12 shows that even when measured under conditions of divided attention, RT can not be used to predict perceived AV temporal order for any of the four observers. As in Section 4.3.4, a clear dissociation is found for all observers, PSS not being equal to the difference between RTs. However, the magnitude of the dissociation (as defined by the difference between predicted and measured PSS) is significantly reduced from that found in Section 4.3.4 (calculated from simple RTs) for all four observers ($F_{3,7} = 20.47$; $p < 0.05$).

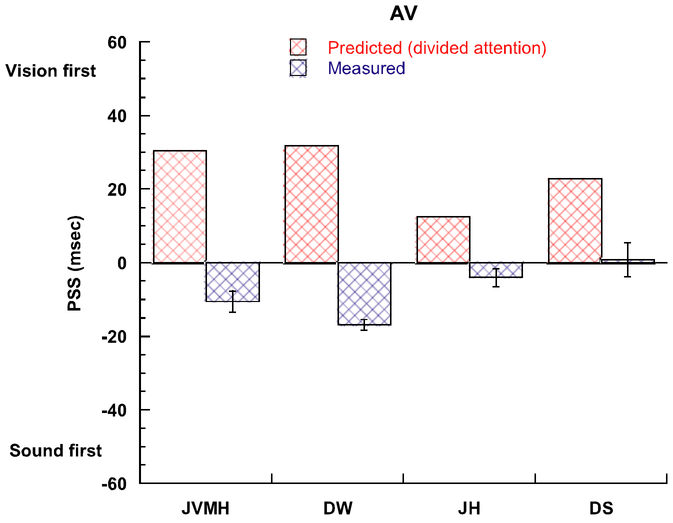


**Figure 4.12.** Comparison of AV PSS values predicted from mean RTs measured under conditions of divided attention (red bars) to those measured using a TOJ task (blue bars) for each observer. Error bars represent one standard deviation either side of the parameter values.

In the AT condition (Figure 4.13), it is again evident that RTs measured under conditions of divided attention cannot be used to predict perceived temporal order. However, the pattern of results differs slightly from the AV pairing; for observer DW the magnitude of the RT/TOJ dissociation has actually increased over that calculated from simple RT, rather than decreased as in the AV pairing. For DS and JH the magnitude of the dissociation has decreased (as in the AV pairing); for JH, the dissociation is now non-significant at the 95% level. For author JVMH, a dissociation is now present where none was evident previously (Section 4.3.4).

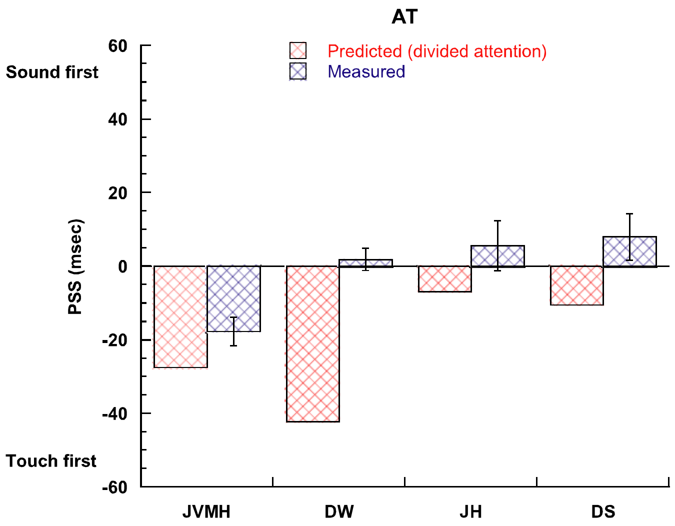


**Figure 4.13.** Comparison of AT PSS values predicted from mean RTs measured under conditions of divided attention (red bars) to those measured using a TOJ task (blue bars) for each observer. Error bars represent one standard deviation either side of the parameter values.

Examination of Figure 4.14 shows that for observer DS, RTs measured under divided attention predict perceived VT temporal order. However, this is not the case for the remaining observers. For JVMH and DW, the magnitude of the dissociation has increased over that calculated from simple RT, as in the AT pairing. For observer JH, conversely, the magnitude of the dissociation has decreased, as in the AV and AT pairings.

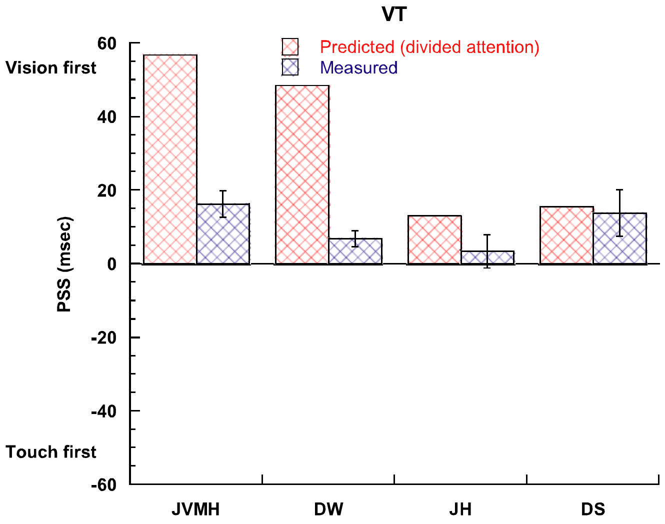


**Figure 4.14.** Comparison of VT PSS values predicted from mean RTs measured under conditions of divided attention (red bars) to those measured using a TOJ task (blue bars) for each observer. Error bars represent one standard deviation either side of the parameter values.

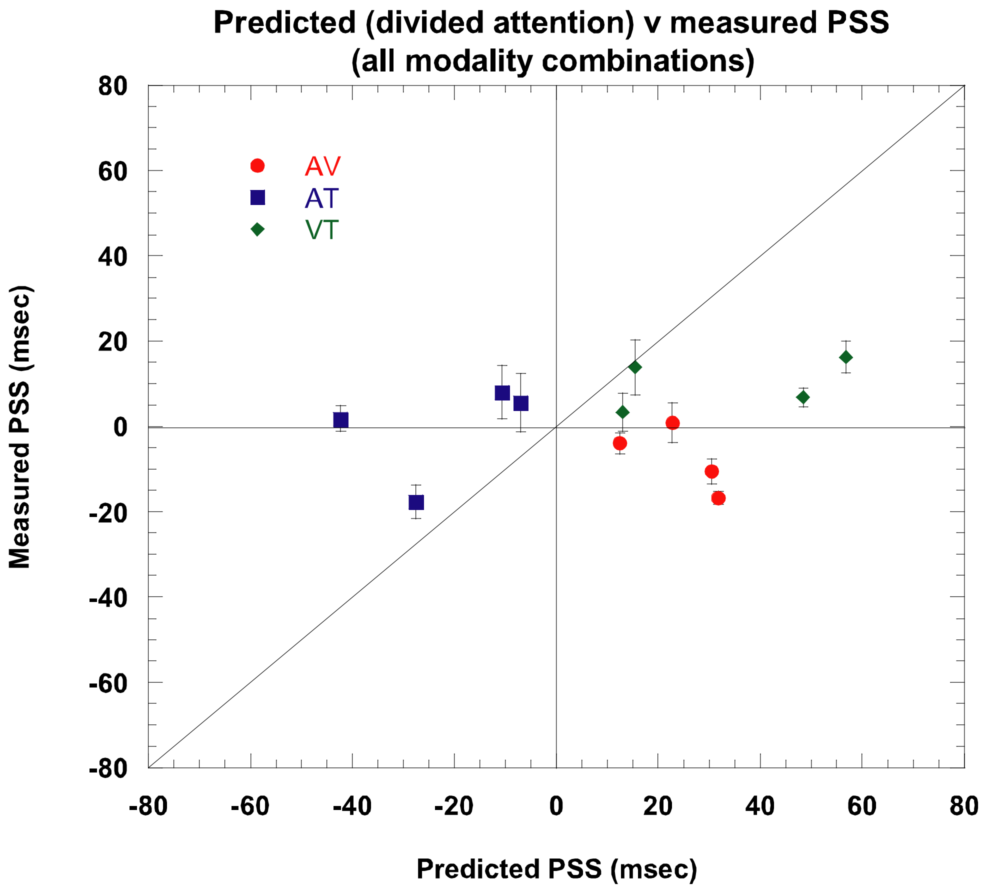


**Figure 4.15.** Scatter plot showing predicted vs. measured PSS for all modality combinations and all observers. The X-axis shows predicted PSS, as given by the algebraic RT differences for each modality pairing (measured under conditions of divided attention). The Y-axis shows measured PSS, as recorded in Section 4.3.4. AV values are shown by red circles, AT by blue squares, and VT by green diamonds. If RT measured under conditions of divided attention is able to predict PSS, values should lie on or around the diagonal line (gradient = 1), where predicted PSS is equal to measured PSS. If the perceptual system fully compensates for differential latencies and RT to maintain a veridical percept of the environment, then values should lie on or around the central horizontal line, where measured PSS = 0. Error bars represent one standard deviation either side of the parameter values.

Thus, it is evident that there are discrepancies between PSS predicted from RT measured under conditions of divided attention and that measured using a TOJ task. This can be visualised in the form of a scatter plot. Figure 4.15 plots PSS predicted from RT measured under conditions of divided attention against that measured empirically and presented in Section 4.3.4. If RT measured in such conditions can be used to accurately predict PSS, then the values should lie on or around the diagonal line (gradient = 1). Only three points from twelve are in such a position (at the 95% confidence limits). The only firm conclusion that can be drawn from this pattern of results is that RT measured under conditions of divided attention cannot be used to reliably predict perceived temporal order in the AV, AT and VT modality pairings. This was confirmed by analysing the correlation between predicted and measured PSS in the AV, AT and VT pairings; the results of this analysis are shown in Table 4.2.

| **Sensory pairing** | **Linear correlation coefficient** | **p** |
|---|---|---|
| **AV** | -0.70 | 0.30 |
| **AT** | 0.41 | 0.59 |
| **VT** | 0.41 | 0.59 |

**Table 4.2.** Correlation coefficient and p-values for PSS predicted by RT measured under conditions of divided attention, and measured using a TOJ task.

The data for each sensory modality were analysed with a two-way repeated measures ANOVA; the factors considered were trial type ('same' or 'different', relative to the modality of the preceding trial) and modality combination (e.g., for the T modality the modality combinations were TV, TA and T (trimodal)). For the visual modality, no significant differences between RTs obtained from 'same' and 'different' trials were recorded ($F_{1,\ 3} = 4.61$, $p > 0.05$). This outcome was also observed in the auditory ($F_{1,\ 3} = 1.15$, $p > 0.05$) and tactile ($F_{1,\ 3} = 0.245$, $p > 0.05$) modalities. In other words, no MSE was observed with the stimuli and experimental parameters employed in the present study. The effect of modality combination was also not significant in the visual ($F_{2, 6}$ = 3.78, $p > 0.05$), auditory ($F_{2, 6}$ = 0.714, $p > 0.05$) and tactile ($F_{2, 6}$ = 1.51, $p > 0.05$) modalities; in other words, simple RT measured under conditions of divided attention for any given modality did not depend on which other modality/modalities observers had to attend to. No significant interaction was found between the two factors ($p > 0.05$).

### 4.4.5 Discussion

After finding the dissociation between simple RT and TOJ in Section 4.3.4, the present experiment aimed to ascertain whether this dissociation can be explained by differing attentional requirements between simple RT and TOJ tasks. The results of this experiment clearly indicate that this is not the case. Although the general trend is for the magnitude of the RT/TOJ dissociation to be *reduced* when RT is measured under conditions of divided attention (a trend particularly evident in the AV condition), a minority of data sets (e.g., JVMH, VT condition) actually show an *increased* dissociation.

A finding of note was that dividing attention between sensory modalities had no significant effect on tactile RT for any of the four observers (conversely, visual and auditory RTs were both slower when attention was divided). Qualitatively, this finding appears similar to the result of a control experiment in an earlier study (Boulter, 1977); however this aspect of the earlier study was not statistically analysed and it is uncertain whether any significant intermodal differences arose.  In another study, using a choice RT task, the tactile modality was shown to be least affected by modality uncertainty (Spence et al., 2001a), in agreement with the present work.

This apparent advantage of tactile processing is likely to have clear ecological advantages: for example, consider an animal navigating an unfamiliar and potentially hostile environment.  The goal of survival clearly requires that any sudden external event is processed by the nervous system as rapidly as possible so that appropriate action may be taken.  In this ecological context, it is likely that tactile events are of the most urgent interest to an organism as, by definition, the cause of such an event is *always* in physical contact with the organism and is potentially an immediate threat.  In comparison, distant events will *always* stimulate the visual and/or auditory systems first, allowing the animal more time to execute an appropriate response to such an event. Thus, the fact that tactile RT is unaffected by division of attention (whilst visual and auditory RTs are slowed) may represent preferential processing of tactile events under conditions of divided attention.  It is also consistent with previous hypothesis that tactile processing requires minimal cognitive resources (Gregory, 1967) and/or that tactile events are inherently alerting in nature (Posner, 1978).  The results also support the conclusion of Ho and colleagues, using a different task to that employed here, that vibrotactile events constitute more effective warning signals than either visual or auditory events (Ho, Spence, & Tan, 2005).

RT results were analysed according to the modality of the preceding trial, i.e. whether the previous trial was presented in the same or a different modality, as previous work suggests that choice RT response latency (Spence et al., 2001a) is slowed by shifting attention between the three sensory modalities. This cost to performance of switching modality has been termed the *modality shifting effect* (MSE) or *exogenous attentional capture* (Turatto et al., 2002). As unconsciously shifting attention away from the tactile modality results in a greater RT cost/MSE than from either the visual or auditory modality (Spence et al., 2001a), it can be deduced that asymmetry exists in the MSE. Given the effects of attention on sensory latency, e.g. (Shore et al., 2001; Spence et al., 2001b; Titchener, 1908; Zampini et al., 2005c), it was important that the effect of this variable was explored. However, no MSE was observed in the current experiment (as defined by a significant difference between the 'same' and 'different' RT values). This finding is compatible with previous work suggesting that a significant MSE measured using an ISI of 150msec was not recorded with an ISI of 650msec (Turatto et al., 2002). As a (random) ISI of 250-750msec was used in the present study, it is likely that any MSE no longer existed at such ISIs as observers had sufficient time to redistribute their attention equally between the modalities.

Examining the reaction time data measured under conditions of modality uncertainty, it appears that those obtained in the present study (on average, approximately 210msec (V), 190msec (A) and 180msec (T)) are faster than those recorded in other studies. For example, previous work has measured auditory RTs (varying with task) of between 420-543msec (Ward et al., 2000), 297-318msec (Colavita, 1974) and 518-636msec (Koppen & Spence, 2007c). It is notable that these studies employed a choice RT task (unlike the simple RT task used in the current experiment for ease of comparison with the simple RTs recorded in Section 4.2.4), and when using such a task the extra cognitive demand of observers causes a slowing of RT. However, another previous study used a simple RT task in conjunction with a very similar method to the present study and still obtained RTs significantly longer than those recorded in the present experiment (Boulter, 1977). This is likely to be ascribable to the use of untrained and unpractised observers and RT values derived from as little as five repetitions per condition (therefore giving observers little opportunity for improvement from their initial untrained RT values).

Returning to the issue of choice RT, closer examination of the literature suggests that the use of such a task may have a profound affect on the *relationships* between V, A and T latency, as well as absolute latency values. For example, in contrast to simple RT, visual and auditory choice RT values have been shown to be indistinguishable (Colavita, 1974), despite the clear transduction and neural latency advantage afforded to audition (Sections 1.1.1, 1.1.2, 2.1.1 and 4.2.4). The relationships between choice RT values in the AT (Occelli, Hartcher-O'Brien, Spence, & Zampini, 2008a) and VT (Hartcher-O'Brien et al., 2008) domains are similarly unrepresentative of the relationships between physiological and neural latencies discussed previously. In addition, a choice RT task requires a discriminative judgment of an observer (e.g., ‘which modality was presented?’) rather than merely stimulus detection.  A further similarity between choice RT and TOJ tasks is that attention is divided between the modalities.  It is therefore possible that as choice RT values appear to be unrelated to transduction and neural latencies, and a discriminative judgment is required whilst attention is divided between the sensory modalities (as in a TOJ task), choice RT may be able to predict PSS.  This possibility is currently untested in the literature.  To test this hypothesis, a choice RT task was devised.

## 4.5 Measurement of Choice Reaction Time

### 4.5.1 Subjects

Author JVMH, DW, JH, AA and CV acted as observers.  Both AA and CV were naïve as to the purpose of the study.  All observers were experienced at a choice RT task.

### 4.5.2 Methods and Stimuli

Visual, auditory and tactile stimuli were the same as used in the previous experiments.  All stimuli were controlled by custom-written software run in MatLab (Mathworks, U.S.A.), which also compensated for the 25msec delay in registering observers' responses caused by the computer hardware and operating system.  AV, AT and VT modality pairings were tested.

### 4.5.3 Procedures

The subject sat in the same darkened room used for the previous experiments. Each experimental run consisted of 85 presentations (40 per modality), with the first five presentations being treated as practise and not included in data analysis. At the beginning of each run, observers were instructed which two modalities would be presented, and which of two response keys corresponded to which modality. They were also told to distribute attention equally between the modalities in question. Each modality was presented an equal number of times within an experimental run, the precise order being generated at random. Each presentation consisted of a single stimulus, to which the observer had to make a speeded response by pressing the appropriate response key. It was emphasised to observers to respond as quickly as possible without sacrificing accuracy. The ISI varied randomly (with a uniform probability) between 250 and 750msec. As with the previous experiments (Sections 4.2.3 and 4.4.3), RTs shorter than 100msec were excluded from data analysis.

After an initial practice run (data not analysed) in each condition, each subject completed five experimental runs in each condition (AV, AT and VT), making a total of 1200 valid presentations per observer (40 presentations * 5 runs * 2 conditions per modality * 3 modalities). As with the previous experiments, the three stimulus pairings (AV, AT and VT) were tested sequentially (e.g., one run of AV, followed by one of AT, one of VT, and so on. After data collection was completed, predicted PSS values were calculated for each observer in each modality pairing as the algebraic difference between choice RT values. Observers AA and CV also performed the crossmodal TOJ task in an identical manner to that described previously (Section 4.3.3) to obtain PSS values for comparison; both observers had previous experience of a TOJ task.

### 4.5.4 Results

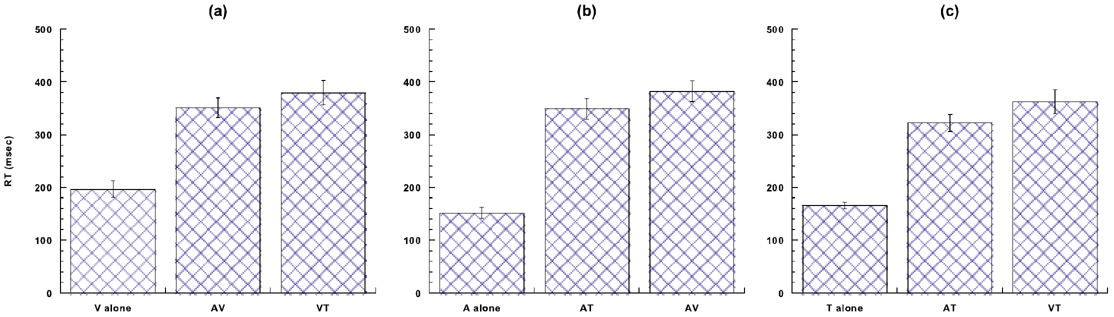


**Figure 4.16(a-c).** Visual (a), auditory (b) and tactile (c) choice RT values for observer JVMH. Each graph shows unimodal RT for that modality (left column) and choice RT for that modality in its two bimodal combinations. Error bars represent one standard deviation either side of the parameter values.

Choice RT values for observer JVMH are shown in Figure 4.16(a-c, corresponding to the visual, auditory and tactile modalities respectively), along with unimodal RT values for comparison. It is immediately obvious that the addition of a response choice increases RT significantly relative to simple RT.

The algebraic differences between choice RTs for the modalities in each pairing (e.g., AV) were used to predict PSS. AV, AT and VT PSS values predicted in this manner are compared with measured PSS values for each observer in Figures 4.17, 4.18 and 4.19, respectively. Examination of the figures shows that, in general, predicted PSS values differ substantially from measured PSS values. This suggests that choice RT is unable to predict PSS. However, the pattern of results appears to differ between observers; most notably, for observer JH predicted PSS values are within the 95% confidence limits of measured PSS values in all three modality pairings. Conversely, for some observers and conditions, choice RT and PSS appear completely unrelated. For example, for observer AA in the AV pairing differences in choice RT values suggest that the observer would require a visual stimulus to lead an auditory stimulus by 5msec for perceived simultaneity, whereas in fact an *auditory* stimulus must lead the visual by 29msec.

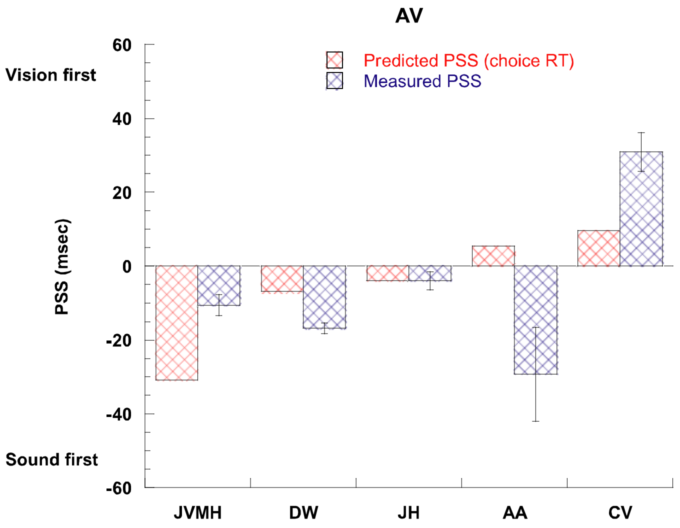


**Figure 4.17.** Comparison of AV PSS values predicted by differential choice RT values (red bars) with those measured using a TOJ task (blue bars) for five observers. Error bars represent one SD either side of the parameter values.

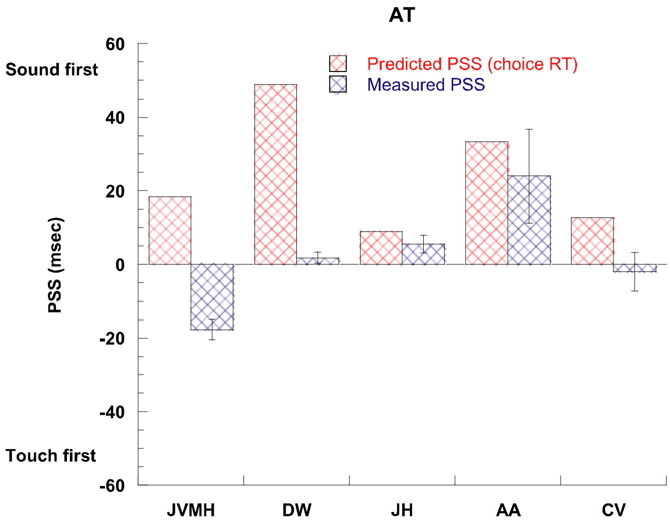


**Figure 4.18.** Comparison of AT PSS values predicted by choice RT (red bars) with those measured using a TOJ task (blue bars) for five observers. Error bars represent one SD either side of the parameter values.

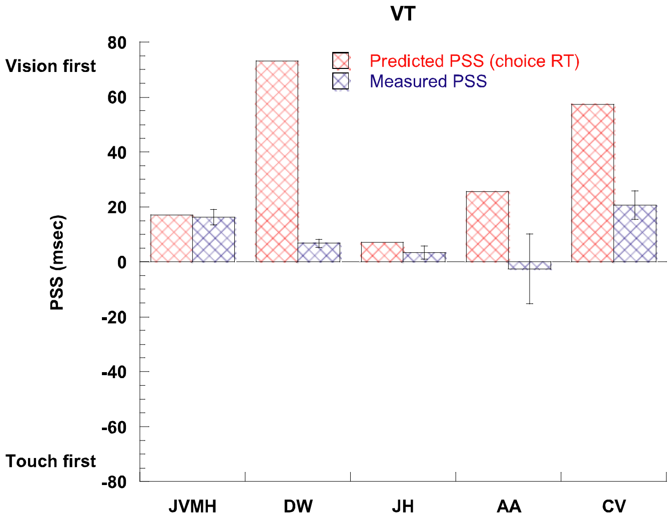


**Figure 4.19.** Comparison of VT PSS values predicted by choice RT (red bars) with those measured using a TOJ task (blue bars) for five observers. Error bars represent one SD either side of the parameter values.

To enable an overview of results in all modality pairings simultaneously, as in Figures 4.6 and 4.15, predicted PSS was plotted against measured PSS. This is shown as Figure 4.20. If choice RT is able to accurately predict PSS, then all data points should lie on the diagonal line where predicted PSS = actual PSS; six data points of 15 satisfy this criteria at the 95% confidence limits. Overall, as with the previous experiments summarised in Figures 4.6 and 4.15, the overall pattern of results is equivocal and the most likely conclusion to be drawn is that choice RT is unable to predict PSS in the AV, AT and VT modality pairings. This was confirmed by analysing the correlation between predicted and measured PSS in the AV, AT and VT pairings; the results of this analysis are shown in Table 4.3.

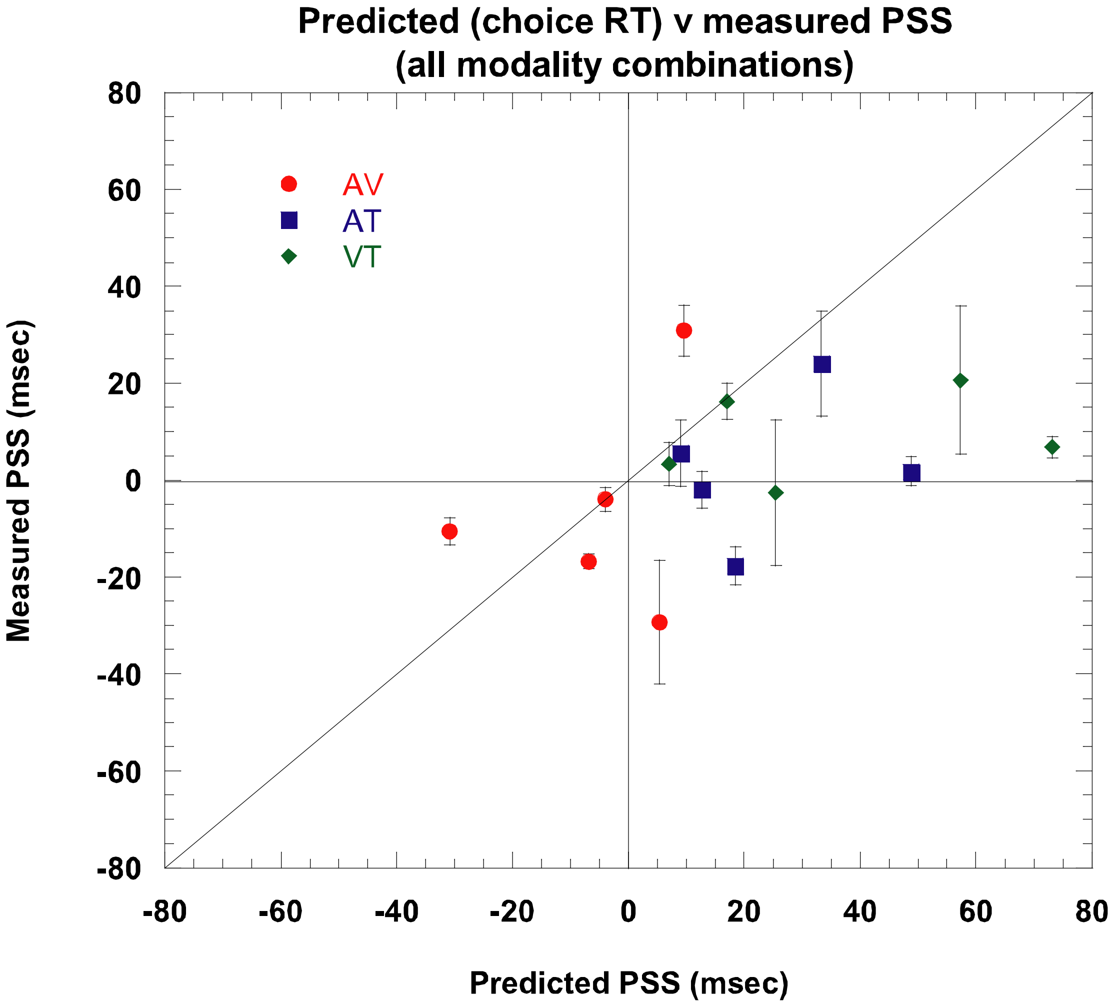


**Figure 4.20.** Comparison of PSS predicted by choice RT with that measured by a TOJ task in the AV (red circles), AT (blue squares) and VT (green diamonds) pairings. If choice RT is able to accurately predict PSS, then the data points should lie on the diagonal line (gradient = 1). If the perceptual system fully compensates for differential latencies and RTs to maintain a veridical percept of the environment, then data points should lie on or around the central horizontal line, where measured PSS = 0. Error bars represent one standard deviation either side of the parameter values.

| Sensory pairing | Linear correlation coefficient | p |
|---|---|---|
| AV | 0.31 | 0.62 |
| AT | 0.30 | 0.62 |
| VT | 0.29 | 0.63 |

**Table 4.3.** Correlation coefficient and p-values for PSS predicted by choice RT and measured using a TOJ task.

### 4.5.5 Discussion

After the previous findings that PSS (as measured using a TOJ task) cannot be predicted by simple RT or RT measured under conditions of divided attention, the possibility that choice RT may be able to predict PSS was investigated. Although this may be the case for individual observers (e.g., observer JH in the present work), overall this hypothesis is not supported by the data. This negative finding suggests that even the addition of a cognitive demand (i.e., the response choice) and the division of attention between modalities cannot match relative perceptual latencies as measured using choice RT with those measured using a TOJ task.

As in previous studies (e.g., Colavita (1974)), the relationships between choice RT values in the present study tend not to reflect the physiological and neural factors discussed in Sections 1.1.1-1.1.3, or the simple RT values presented in Figure 4.1. For example, examination of Figure 4.17 shows that for all five observers, tactile choice RT is shorter than auditory (as evidenced by the predicted requirement of an auditory lead for perceived simultaneity in a TOJ task), despite the neural latency advantage afforded to audition. The cognitive process of identifying the modality of each presentation and executing the appropriate response appears to strongly influence the relationships between choice RT values. Absolute RT is also affected, as illustrated by Figure 4.16; specifically, choice RT appears significantly longer than simple RT. This result was expected on the basis of previous literature showing such effects (e.g., Colavita (1974)), and was not analysed further as the relationship between choice RT in the different modalities was the primary parameter of interest.

Taken as a whole, the experiments presented in this chapter clearly suggest that RT and multimodal TOJ measure some fundamentally different aspect(s) of sensory latency; RT to stimuli in different modalities cannot predict PSS, and vice versa. This dissociation between the two measures of sensory latency appears robust to manipulations of attentional and cognitive demands in the RT task. Although the measurement of RT under various conditions and subsequent comparison to multimodal PSS in the AV, AT and VT pairings presented here is novel, a variety of other dissociations (in addition to sensory modality) between these measures of sensory latency have been previously observed. These include spatial frequency (Tappe, Niepel, & Neumann, 1994), stimulus duration (Jaskowski, 1992), stimulus rise time (Jaskowski, 1993), and intensity (Roufs, 1974; Sanford, 1974). The results of such studies typically show that changes to stimulus parameters (e.g., intensity) have a significantly larger affect on RT than on PSS. A variety of models and hypotheses have been proposed in order to explain these dissociations.

The simplest of these models (referred to elsewhere as the *canonical model* (Miller & Schwarz, 2006)) suggests that RT and TOJ are mediated by the same internal response and decision process, and that both are dependent on the duration of a shared perceptual detection process (Gibbon & Rutschmann, 1969). In other words, both RT and TOJ are dependent purely on the timescale of stimulus detection. If this were the case, we would expect RT and TOJ to correlate as measures of sensory latency: the data presented here clearly demonstrate that this is not the case. We would also expect changes in stimulus parameters to affect RT and PSS equally, which is also inconsistent with other experimental findings (Cardoso-Leite et al., 2007; Jaskowski, 1992, 1993; Roufs, 1974; Sanford, 1974; Tappe et al., 1994).

Other workers have proposed that the nervous system uses different temporal information about the sensory events depending on whether a RT or TOJ task is employed (Sternberg & Knoll, 1973). Specifically, in a RT task the importance of response speed is emphasised, observers are therefore likely to respond to a signal as soon as the elicited neural activity reaches a level corresponding to their response criterion: in other words, respond to the perceived *onset* of the signal. Conversely, as a TOJ task is unspeeded, observers do not need to respond as quickly as possible and may therefore base their responses on a comparison of the relative timing of the *peaks* of the internal responses evoked by stimulus presentation (Sternberg & Knoll, 1973). Therefore, the same decisional mechanism is employed at different times upon the same internal response. Whilst the reasoning behind this theory (which has been termed the *hypothesis of different time markers* (Jaskowski, 1996)) is credible, it is difficult to reconcile with the results obtained here. In particular, the stimuli used in the present work are too short in duration (10msec) for such a strategy to be implemented consciously; however, the wide range of PSS values between observers is more consistent with a perceptual/cognitive decisional mechanism. It is also unclear how the timing of the peak internal response could be employed as a time marker if observers were forced to choose the order of stimulus *offset* (rather than onset) of extended 2000msec (rather than transient) stimuli, as in previous work (Allan, 1975).

The dissociations between RT and TOJ have also been proposed to reflect different response criteria in the two tasks. These *criterion shift* models assume that sensory activation proceeds faster for (e.g.) a more intense stimulus than a less intense one, and that stimulus detection occurs when sensory activation reaches a certain criterion level. The key assumption of these models, which enables them to explain the intensity dissociation between RT and TOJ, is that the response criterion is higher in the RT task than in the TOJ task (Sanford, 1974). This assumption may appear counterintuitive in that as observers' responses are speeded in the RT task, it would be most logical were criterion also lowest in this task (Tappe et al., 1994).

However, Miller and Schwarz (2006) have suggested that such a strategy may in fact optimise response accuracy in a TOJ task; if a high intensity stimulus is first to reach a relatively high criterion even if it occurs after a low intensity stimulus (i.e., if the internal representation of the intense stimulus ‘overtakes’ that of the earlier, less intense stimulus), then observers will erroneously respond that the stimulus occurred first. Conversely, a less intense stimulus occurring before a more intense stimulus is more likely to reach a relatively low detection criterion level, maximising the chances of a correct response (Miller & Schwarz, 2006). However, this model implies that the intensity effects on TOJ should be minimal (if present at all), as observers would be likely to simply adopt an arbitrarily low criterion in the TOJ task in order to maximise their correct responses. The model would thus predict an arbitrarily small effect of intensity on TOJs, which is inconsistent with the significant effects routinely observed (Miller & Schwarz, 2006).

Miller and Schwarz therefore developed their *diffusion model* in an attempt to resolve this inherent problem in criterion-shifting models. This model assumes a lower criterion in the TOJ task than an RT task (as in criterion shifting models), but with the same detection process used in both tasks (as in the canonical model). When the two events in the TOJ task have both been detected, the precise onset times are estimated retrospectively and from this estimation the onset order is determined; an observer is assumed to always respond that the stimulus which was detected first was actually presented first. The model is presented in full, with quantitative justifications for each assumption, in Miller and Schwarz (2006).

Although Miller and Schwarz found good fits of their model to previously acquired data, it is difficult to apply to the results presented here. This is because the diffusion model was developed with the *intensity* (rather than modality) dissociation between RT and TOJ as its primary consideration, reflecting the fact that the model is primarily unimodal in conception. It is currently unclear whether unimodal and crossmodal relative timing judgments are quantitatively and/or qualitatively similar, and the stimulus artefacts frequently observed in unimodal TOJ studies (Section 2.1.3) make comparisons between unimodal and bimodal studies problematic. However, the available evidence suggests that unimodal relative temporal judgments are more precise when directly compared to bimodal conditions within the same study (Fujisaki & Nishida, 2007a; Sinex, 1978), possibly reflecting differences in uni- and bimodal processing. It has been proposed that the diffusion model is also able to account for the full range of dissociations between RT and TOJ, including that of sensory modality (Miller & Schwarz, 2006), but this currently remains untested.

Another class of model posits that RT and TOJ may be mediated by different neural mechanisms. For example, one group of researchers (Neumann, Esselmann, & Klotz, 1993; Tappe et al., 1994) has proposed that a RT task does not require conscious awareness of a stimulus, simply an ‘automatic’ response such as the press of a button. Thus, as long as sensory processing can access a route to the motor pathway, a response can be made. Conversely, a TOJ task requires a *conscious* decision as to which stimulus came first (Neumann et al., 1993; Tappe et al., 1994). This hypothesis is partly in keeping with available data. It is known that visual, auditory and tactile RT measurement activates (in addition to modality-specific sensory cortices) a network including right posterior superior temporal cortex, right and left premotor cortex, right occipitotemporal gyrus, and medial frontal gyrus/supplementary motor area (Kansaku et al., 2004). The neural correlates of multimodal TOJ are currently unclear, although one AV study (using a SJ rather than TOJ task) has implicated the insula, cerebellum, inferior frontal gyrus, inferior parietal lobe, superior colliculus and posterior thalamus (Bushara et al., 2001). Therefore, the notion of detection tasks (e.g., RT) and relative timing judgments (e.g., TOJ and SJ) employing different neural circuits is credible.

If the above hypothesis was correct, however, we may expect choice (rather than simple) RT to correspond to PSS as both require a conscious decision of the observer (which stimulus was presented/which was presented first); Figures 4.16 - 4.19 suggest that this is not the case. Also, as commented previously, any results which suggest that RT and TOJ are mediated by separate neural mechanisms may also be consistent with the hypothesis of different time markers (Jaskowski, 1996; Neumann et al., 1993). In other words, a single mechanism using different stimulus timing information in RT and TOJ tasks (the time at which the neural response exceed criterion vs. the timing of the peak neural responses, respectively) may yield similar behavioural results to two distinct mechanisms for RT and TOJ. As stated by Jaskowski (1996), the only significant difference between the hypothesis of different time markers and Neumann et al.'s suggestion of two distinct neural mechanisms is that the latter assumes that the response process in an RT task is unconscious, whereas Sternberg and Knoll (1973) make no explicit assumption about this. In summary, the results presented here do not unambiguously support any of the above models. However, they clearly argue against the canonical model (Gibbon & Rutschmann, 1969) whilst being consistent with the notion that RT and TOJ are subserved by different decisional mechanisms and/or internal information.

To conclude, the data presented here show that sensory latency as measured by simple and choice RT, as well as simple RT whilst attention is divided between the sensory modalities, does not predict PSS as measured with a TOJ task. They are also not consistent with any of the quantitative models discussed above which attempt to account for similar dissociations. However, the data are consistent with the notion that RT and TOJ may measure some fundamentally differing aspect(s) of temporal processing. A potentially significant factor is that the models discussed above were primarily formulated and tested with reference to unimodal visual processing, and whether multisensory comparisons (e.g., AT) are qualitatively and/or quantitatively comparable to visual unimodal processing is currently unclear.

# Chapter 5

## 5.1 The effects of asynchrony adaptation on perceived audiovisual, audiotactile and visuotactile temporal order

### 5.1.1 Introduction

At least two physical factors influence the perceived timing of distal multisensory events. Firstly, the differential velocities of light and sound in air ensure that distant auditory stimuli arrive progressively later than their visual counterparts (Section 2.1.3). Secondly, differential transduction and conduction latencies (Sections 1.1.1-1.1.3) ensure that visual, auditory and tactile stimuli can arrive synchronously at their respective receptor surfaces, yet arrive asynchronously at their respective primary sensory cortices. As such, it could be considered surprising that the auditory, visual and tactile components of a single event are usually perceived as simultaneous.

The fact that, under most real-world situations, perceptual synchrony is the norm itself suggests that the nervous system possesses a degree of temporal elasticity, which allows it to recalibrate signals in different sensory modalities to maintain a veridical, synchronous percept of the world. In recent years, this concept of temporal sensory recalibration has been the subject of mounting scientific interest (Arnold et al., 2005; Fujisaki et al., 2004; Harrar & Harris, 2005; Heron et al., 2007; Keetels & Vroomen, 2007; Kopinska & Harris, 2004;

Lewald & Guski, 2004; Navarra et al., 2007; Sugita & Suzuki, 2003; Vroomen et al., 2004). Particular interest has been focussed on a proposed perceptual mechanism to maintain perceived audiovisual (AV) synchrony over a range of observer-source distances in spite of the significantly different propagation velocities of sound and light in air; a review of the important findings and controversies relating to this body of work is presented in Section 2.1.3.

A novel mechanism for temporal sensory recalibration has been described by both Fujisaki et al. (2004) and Vroomen et al. (2004). Both studies demonstrated shifts in the Point of Subjective Simultaneity (PSS - the *physical* temporal offset between two stimuli required for *perceptual* simultaneity) following exposure to a series of asynchronous AV stimulus pairs. Specifically, following a period of exposure to asynchronous pairs (e.g. vision leads sound), subjects required the same polarity of asynchrony for the two to be perceived as simultaneous (e.g. vision must now *physically* lead sound for *perceptual* simultaneity). This temporal recalibration was recorded with the use of both Temporal Order Judgment (TOJ; Section 3.3.1) and simultaneity judgment (SJ; Section 3.3.2) tasks. The typical pattern of responses made by observers in the TOJ task is illustrated in Figure 5.1.

On the basis that only distant AV events suffer the aforementioned temporal misalignment (vision physically leads sound at significant distances), it has been suggested that this type of ‘lag adaptation’ mechanism is likely to be restricted to the AV domain (Miyazaki, Yamamoto, Uchida, & Kitazawa, 2006). By definition, a tactile stimulus must be in direct contact with an observer.

Thus, there appears to be no obvious ecological advantage in possessing a lag adaptation mechanism that involves this modality (Miyazaki et al., 2006).

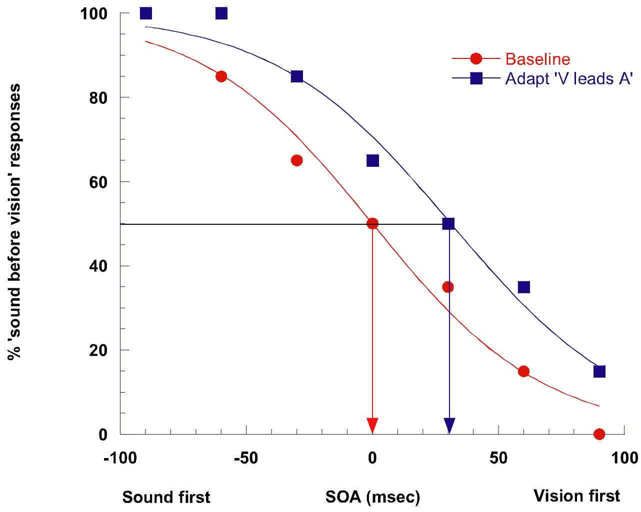


**Figure 5.1.** Psychometric functions derived from hypothetical data illustrating observers' responses in an AV TOJ task before and after exposure to asynchronous stimulus pairs (vision leads sound). PSS corresponds to 50% 'sound before vision' responses on the y-axis. Prior to adaptation ('baseline' condition, red circles), the observer has a PSS of approximately zero (i.e., *perceptual* simultaneity approximates to *physical* simultaneity). Following exposure to 'vision leads sound' asynchronous stimulus pairs (blue squares), the observer is now more likely to respond 'sound before vision' at each stimulus-onset asynchrony (SOA). The effect of this altered pattern of responses is that the psychometric function is shifted laterally, such that chance performance (and therefore PSS) now corresponds to a physical lead of vision over sound.

A recent investigation of asynchrony adaptation in the audiotactile (AT) pairing appears to confirm the hypothesis of Miyazaki and colleagues. Navarra et al. (2007) found that exposure to asynchronous 'sound leads touch' AT stimulus pairs resulted in PSS values similar to those obtained following exposure to synchronous AT stimulus pairs. Sensitivity to asynchrony, however, was reduced (as evidenced by higher Just-Noticeable Difference (JND) values) following asynchrony adaptation relative to the 'adapt synchronous stimulus pairs' condition. The results of this study (Navarra et al., 2007), and earlier work in the AV pairing (Fujisaki et al., 2004; Vroomen et al., 2004) suggest that perception of crossmodal temporal order in the AV pairing is influenced by recent experience, whilst this is not the case in the AT pairing.  Given the compelling argument that temporal recalibration in the tactile modality is unnecessary and even undesirable (Miyazaki et al., 2006), this pattern of results appears credible; however, this argument fails to account for the earlier finding that recalibration readily occurs following exposure to a perceptually obvious lead of sound over vision (Fujisaki et al., 2004; Vroomen et al., 2004), despite the impossibility of such a situation in a natural environment.  This suggests that temporal flexibility may not be directly related to long-term exposure to ‘vision leads sound’, thereby introducing the possibility of adaptive timing shifts involving the tactile modality.

The most obvious way to investigate the hypothesis of Miyazaki and colleagues (2006) would be to undertake an investigation into asynchrony adaptation in the visuotactile (VT) pairing, as no data relating to the effects of asynchrony adaptation on perceived simultaneity in this stimulus pairing is currently in the public domain.  Should no adaptive shifts in perceived simultaneity be observed, the hypothesis of Miyazaki et al. (2006) would be confirmed.  If, on the other hand, PSS shifts following asynchrony adaptation are observed in the VT pairing, this would be difficult to reconcile with previous work in the AT pairing (Navarra et al., 2007).  This is because it is not immediately apparent why only one or the other of the AT and VT pairings would display adaptive PSS shifts similar to those in the AV pairing.  As there would appear to be no need for a mechanism mediating adaptive temporal recalibration involving tactile stimuli (Miyazaki et al., 2006), a more intuitive pattern of results would feature *neither* or *both* of the AT and VT pairings demonstrating recalibration of perceived simultaneity following asynchrony adaptation.

With these factors in mind, an experiment was devised to investigate the effects of adaptation to asynchronous stimulus pairs on perceived temporal order in the AV, AT and VT pairings. This would allow comparison of the effects of asynchrony adaptation on PSS and JND in the three modality pairings using identical stimuli and methods in each pairing. It was hoped that the data would shed light on the nature of the neural mechanism(s) underpinning temporal recalibration in the sensory pairings tested.

With regard to the possibility of adaptive shifts in PSS, it was likely that one of at least two possible outcomes would be found: firstly, that perceived simultaneity would be altered by adaptation in the AV pairing but not in the AT or VT pairings. Such a result would be consistent with previous data from the AV pairing (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Miyazaki et al., 2006; Vroomen et al., 2004), AT pairing (Navarra et al., 2007), and previous hypothesis (Miyazaki et al., 2006). The implication of such a result would be that no mechanism exists to recalibrate the perceived timing of tactile stimuli. Any temporal recalibration in the AV pairing would therefore suggest the existence of a neural mechanism serving that pairing alone.

A second possibility is that adaptive PSS shifts would be found in all three pairings. This possibility arises due to the fact that recalibration effects are found following exposure to suprathreshold ‘sound leads vision’ stimulus pairs (Fujisaki et al., 2004; Vroomen et al., 2004), despite a perceptually obvious lead of sound over vision being impossible in a natural environment.  It is therefore possible that asynchrony adaptation is not simply dependent on a lifetimes’ experience of vision leading sound with increasing distance.  Such a finding would be consistent with a single neural mechanism mediating sensory temporal recalibration in all three sensory pairings tested.  This hypothesis would be given even greater weight if the size of any temporal recalibration effects were found to be similar across the three sensory pairings.  In these circumstances, it would be logical to assume the existence of a single, relatively late-stage neural mechanism recalibrating *all* sensory input with respect to recent experience.

Consistent with the idea of a late-stage neural mechanism for temporal recalibration, previous work provides evidence that the recalibration mechanism in the AV pairing is not dependent on low-level stimulus properties being identical in adapt and test phases.  Fujisaki et al. (2004) showed that exposure to simple flash/click asynchronous AV stimuli modulated perception of simultaneity as measured with the bounce/stream task[7] (Sekuler, Sekuler, & Lau, 1997).  Alteration of the laterality of the auditory stimulus (right vs. left ear headphone presentation) between adapt and test phase also did not affect the temporal recalibration process (Fujisaki et al., 2004).  In other

---

[7] The bounce-stream illusion is a bistable visual display featuring two stimuli approaching each other and travelling through each other; observers can perceive the two stimuli as either ‘bouncing’ off each other, or ‘streaming’ through each other (Sekuler et al., 1997).  The perception of ‘bouncing’ may be induced by a tone presented around the moment the two stimuli spatially co-incide.  The centroid of an observers’ bounce response distribution provides an estimate of PSS (Fujisaki et al, 2004).  Thus, the task does not require an explicit

words, temporal recalibration was still evident when the stimuli differed substantially between adapt and test phase. This work indicates that, at least in the AV pairing, temporal recalibration following exposure to asynchronous stimulus pairs is unlikely to be mediated at an early stage of the perceptual system; were this the case, it is more likely that temporal recalibration would *not* have been evident when substantially altering low-level stimulus characteristics between adapt and test phases.

Also unclear at present is the effect of asynchrony adaptation on observer sensitivity to asynchrony in the VT pairing. Navarra et al. (2007) showed a reduction in sensitivity following adaptation to asynchronous ‘sound leads touch’ (relative to synchronous) AT stimulus pairs, but the effects of a lead of touch over sound were not investigated. Conversely, an AV study examining the effects of adaptation to a lead of vision over sound found that JND did *not* differ significantly between adapted and unadapted conditions (Heron et al., 2007). Given that this study investigated the role of AV asynchrony adaptation in a ‘real-world’ environment, which necessarily excluded investigation the effects of a lead of sound over vision, it is unknown whether this pattern would remain when considering a lead of sound over vision. Although one study has shown that JND does not systematically vary as a function of AV asynchrony adaptation polarity or relative spatial location of the auditory and visual stimuli, in this case no unadapted or synchronous adaptation condition was included for comparison (Keetels & Vroomen, 2007). Previous studies investigating *both* polarities of AV asynchrony adaptation

---

judgment of simultaneity and therefore differs from a SJ task even though the two tasks both provide an estimate of PSS.

relative to unadapted (Fujisaki et al., 2004) and ‘adapt physically synchronous AV stimuli’ (Fujisaki et al., 2004; Vroomen et al., 2004) conditions have not included any data or analysis relating to sensitivity of TOJs. In summary, thus far the effect of asynchrony adaptation on perception of temporal order and sensitivity to asynchrony in the AV, AT and VT pairings is unclear.

With regard to the potential effects of asynchrony adaptation on JNDs, one of two experimental outcomes was anticipated. Firstly, as observed by Navarra et al (2007) in the AT pairing, JND may increase following adaptation to asynchronous (sound leading touch) stimulus pairs. In the study of Navarra et al., this increase was relative to a measure of sensitivity obtained following adaptation to physically synchronous stimulus pairs. This outcome would be consistent with asynchrony adaptation affecting a dedicated asynchrony detection circuit such as that described previously in the AV pairing (Bushara et al., 2001): in other words, repeated exposure to asynchrony would fatigue the asynchrony detection mechanism and make it more difficult for observers to detect asynchrony of the test stimuli. Alternatively, asynchrony adaptation may have no effect on JND relative to unadapted baseline measures of sensitivity. Such an outcome is suggested by previous work in the AV pairing, which examined the effects of adaptation to a lead of vision over sound (Heron et al., 2007). This would suggest that the process of temporal recalibration is independent of any such asynchrony detection network.

With this in mind, the proposed series of experiments had the following aims:

- to test the hypothesis of Miyazaki et al. (2006) that temporal recalibration following asynchrony adaptation is unique to the AV stimulus pairing

- to assess the effect of asynchrony adaptation on observers sensitivity to temporal order in the AV, AT and VT pairings

## 5.2 Assessment of asynchrony adaptation on perception of crossmodal temporal order

### 5.2.1 Subjects

Observers were author JVMH, DW and JH, and trained observer CV (who was naïve as to the purpose of the experiment). All observers had extensive experience of a TOJ task.

### 5.2.2 Methods & stimuli

The visual, auditory and tactile stimuli have been described previously (Section 4.2.2). Presentation of the stimuli was controlled by custom-written software run in MatLab (Mathworks, U.S.A.) on a Dell desktop PC. The relative timings of all stimuli were verified by simultaneous capture on a multiple trace oscilloscope.

### 5.2.3 Procedures

Baseline measures of PSS were obtained for each observer in each modality pairing (AV, AT and VT) by use of a temporal order judgment (TOJ) task. A trial consisted of pairs of stimuli presented at one of seven physical stimulus onset asynchronies (SOA): -90, -60, -30, 0 (simultaneous), 30, 60 and 90msec, which were randomly interleaved within a method of constant stimuli. In the present study, positive SOAs always refer to a visual lead and negative always to a tactile lead. The interval between stimulus pairs varied randomly (with a uniform probability) between 250-750msec. After each presentation, the observer made an unspeeded TOJ as to 'which modality came first' and responded via the computer keyboard. Each experimental run contained 10 presentations at each SOA and each observer completed five experimental runs in each of the three stimulus pairings, making a total of 1050 presentations per observer (10 presentations * 7 SOAs * 5 experimental runs * 3 modality pairings).

In the adapted conditions, an initial period of adaptation preceded each experimental run; this consisted of 100 pairs of stimuli separated by an SOA of either +90 or -90msec (consistent within an experimental run). This value was chosen because previous work in the AV domain suggested that this asynchrony level would be sufficient to elicit quantifiable adaptive shifts in PSS (Fujisaki et al., 2004; Vroomen et al., 2004). As in the baseline condition, the interval between presentation of stimulus pairs varied randomly (with a uniform probability) between 250 and 750msec. Observers simply attended to the adapting stimulus pairs in the absence of any cognitive task. After the adaptation phase, a pause of one second duration alerted the observer that the adaptation phase was complete and the test phase was imminent. This one second pause was followed by four ‘top-up’ asynchronous (e.g. +90msec) stimulus pairs – identical to those in the adaptation phase – plus a fifth ‘test’ pair with one of seven SOAs separated by 30msec steps. By making their TOJ response to this test pair (in an identical manner as in the baseline, unadapted condition), the next phase of the top-up/test cycle was initiated. The experimental paradigm is illustrated in Figure 5.2.

Pilot data collected by observer DW suggested that significant shifts in PSS (which were in the direction of the adapting asynchrony, as in previous work (Fujisaki et al., 2004; Vroomen et al., 2004) and as illustrated in Figure 5.1) occurred in all three sensory pairings following asynchrony adaptation. Therefore, to centralise the resultant psychometric function and optimise its fit to the data points, the range of offsets tested varied depending on the adapting lag. When adapting to a positive lag (positive always relating to a visual lead), the offsets tested were -60, -30, 0 (i.e., simultaneous), 30, 60, 90 and 120msec; when adapting to a negative lag (negative always referring to a tactile lead), the offsets tested were -120, -90, -60, -30, 0 (i.e., simultaneous), 30 and 60msec. Each experimental run consisted of 10 test presentations at each SOA, and each observer completed five experimental runs following both positive and negative asynchrony adaptation (+90 and -90msec) in each of the three modality pairings. This made a total of 2100 test presentations per observer (10 presentations * 7 SOAs * 5 experimental runs * 2 asynchrony polarities * 3 modality pairings) in the adapted conditions.

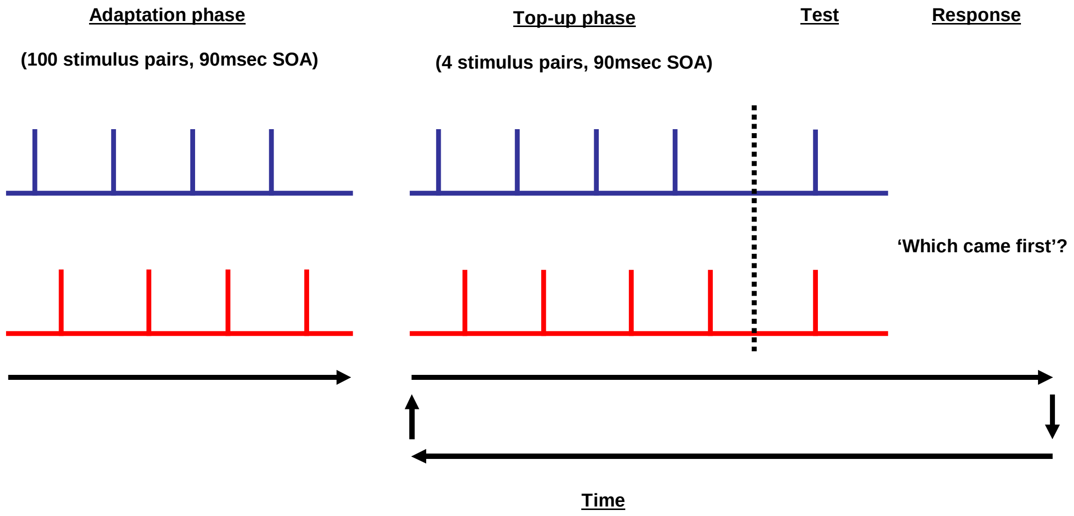


**Figure 5.2.** Schematic showing the paradigm used during investigation of the effect of adaptation to asynchronous stimulus pairs on perception of crossmodal temporal order. Observers were passively exposed to 100 stimulus pairs (with 90msec SOA), with a 250-750msec interval between each stimulus pair ('Adaptation Phase' on figure). After the last stimulus pair, a pause of one second alerted observers that the test phase was about to begin. Each test presentation was preceded by four further asynchronous stimulus pairs ('Top-up Phase' on figure), identical to those in the adaptation phase. The fifth presentation ('Test' on figure) was of two stimuli presented at one of a range of asynchronies, as in the baseline condition. Observers made an unspeeded judgment as to the perceived temporal order of the test stimulus pairing and responded via a computer keyboard. Following the observers' response, the cycle of top-up adaptation followed by test stimuli presentation was repeated.

It is important to note that during data collection, all observers performed experimental runs in all of the conditions in a quasi-random order, with alternation between adapted and non-adapted conditions. This step was taken to minimise the chance that any potential practice-related improvement in sensitivity to temporal order would affect the pattern of results, and was considered important in view of the planned analysis of JND.

For all observers, the percentage of ‘sound-first’ or ‘touch-first’ responses for each condition was plotted as a function of SOA and fitted with a logistic function identical to that described previously (Sections 3.1.3 and 4.3.3).

### 5.2.4 Results

Psychometric functions in the AV modality pairing for observer JVMH can be seen in Figure 5.3. Inspection of this figure reveals a clear lateral separation between the adapted function (red circles/curve and blue squares/curve) and the unadapted baseline function (black triangles/curve). Following adaptation, perceived asynchrony shifts away from baseline toward the adaptation phase’s asynchrony value. For example, adaptation to +90msec of AV asynchrony (vision leads sound) increases the post-adaptation proportion of ‘sound-first’ responses (Figure 5.3 – red circles) at all SOA levels, consistent with a negative aftereffect. The reverse pattern of results can be observed for the -90msec AV adaptation condition (Fig 5.3 – blue squares). For each function, arrows indicate the relevant PSS on the *x*-axis. It can clearly be seen that AV PSS has been shifted by adaptation to asynchronous AV stimulus pairs. Inspection of the slope of the psychometric functions suggests that observer sensitivity to AV asynchrony is unaffected by the recalibration process, as evidenced by the similarity between the slope of the unadapted baseline function (black curve) and both adapted functions (red and blue curves).

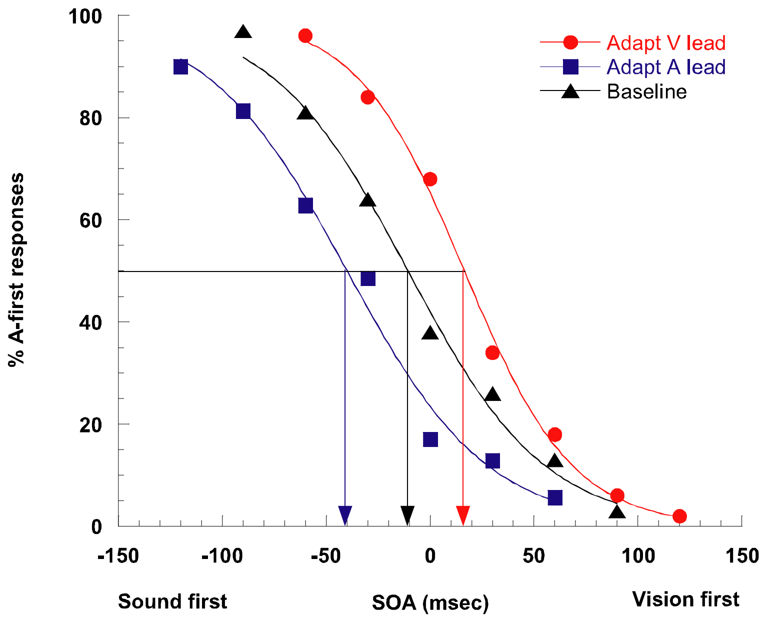


**Figure 5.3.** Psychometric functions for observer JVMH in the AV modality pairing, derived from baseline (unadapted) TOJs (black triangles/curve), and TOJs made following a period of exposure to asynchronous stimulus pairs (red circles/curve and blue squares/curve). The red circles/curve show data obtained following exposure to a lead of vision over sound, whilst the blue squares/curve show data obtained following exposure to a lead of sound over vision. Chance performance is indicated by the horizontal black line on all graphs, and the SOA corresponding to this value (the PSS) is indicated by the arrows between individual functions and the x-axis.

Figures 5.4 and 5.5 show psychometric functions relating to the AT (Figure 5.4) and VT (Figure 5.5) modality pairings for observer JVMH. Inspection of these figures shows essentially the same pattern of results in the AT and VT pairings as in the AV pairing: exposure to asynchronous stimulus pairs shifts the PSS in the direction of the adapting asynchrony. For example, Figure 5.4 shows that exposure to ‘touch leads sound’ stimulus pairs (blue squares/curve) has shifted the PSS, such that observer JVMH now requires an increased *physical* lead of touch over sound for *perceptual* simultaneity.

As in the AV pairing, inspection of the slopes of the psychometric functions suggests that the process of adaptation has not affected the sensitivity of this observer to AT or VT asynchrony.

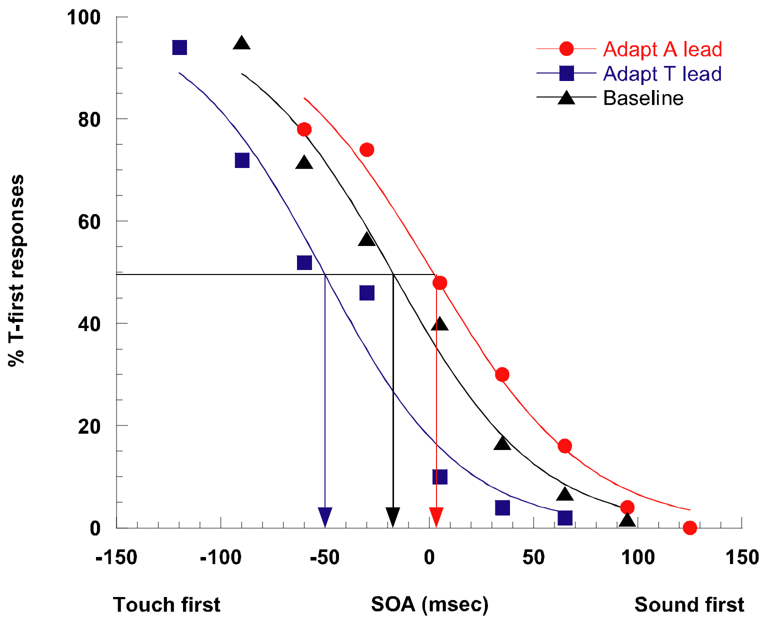


**Figure 5.4.** Psychometric functions for observer JVMH in the AT modality pairing, derived from baseline (unadapted) TOJs (black triangles/curve), and TOJs made following a period of exposure to asynchronous stimulus pairs (red circles/curve and blue squares/curve). The red circles/curve show data obtained following exposure to a lead of sound over touch, whilst the blue squares/curve show data obtained following exposure to a lead of touch over sound. PSS is indicated by the arrows between individual functions and the x-axis.

Figure 5.6 shows the shifts in PSS (relative to baseline) for the three modality pairings averaged across all four observers. The pattern of results is consistent for all conditions – perceived timing is strongly influenced by recent experience, irrespective of the sensory pairing or the polarity of the adapting asynchrony. Adapted PSS shifts for all observers were combined within a 2-way repeated measures ANOVA which revealed that the mean PSS shift was not significantly different across the three modality pairings ($F_{2,\ 6} = 2.47$, $p > 0.05$), but that the effect of adaptation polarity within each modality pairing was highly significant ($F_{1,\ 3} = 53.08$, $p < 0.01$). There was no significant interaction between these two factors ($F_{2,\ 6} = 2.13$, $p > 0.05$), indicating that the effect of polarity was consistent across each modality pairing. In other words, adaptation to positive asynchrony causes a significant positive shift in PSS in all three modality pairings, whilst adaptation to negative asynchrony causes a significant negative shift in PSS.

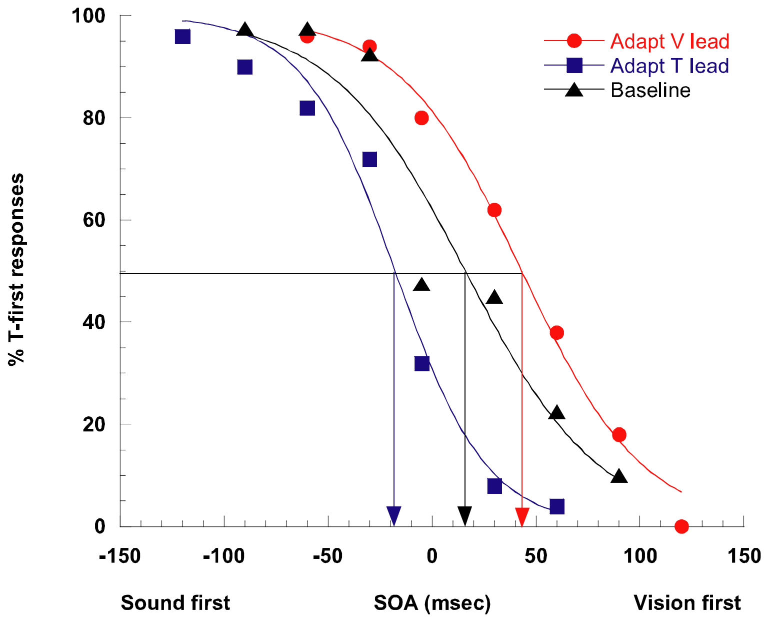


**Figure 5.5.** Psychometric functions for observer JVMH in the VT modality pairing, derived from baseline (unadapted) TOJs (black triangles/curve), and TOJs made following a period of exposure to asynchronous stimulus pairs (red circles/curve and blue squares/curve). The red circles/curve show data obtained following exposure to a lead of vision over touch, whilst the blue squares/curve show data obtained following exposure to a lead of touch over vision. PSS is indicated by the arrows between individual functions and the x-axis.

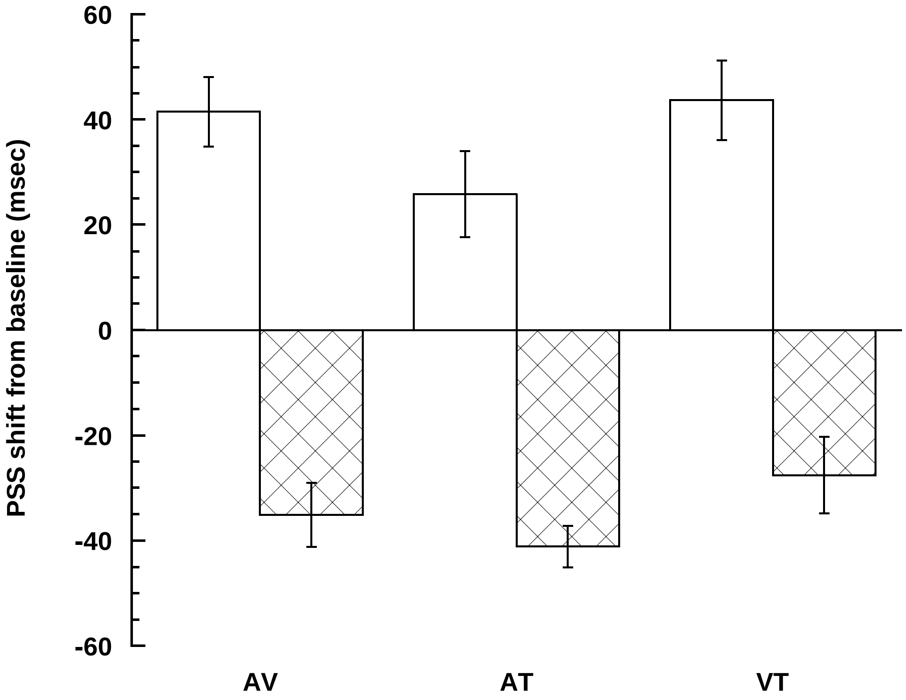


**Figure 5.6.** Graph showing mean PSS shifts relative to baseline following adaptation to positive (unfilled bars) and negative (hatched bars) in each of the three modality pairings. Positive asynchrony always refers to a visual lead, whilst negative asynchrony relates to a tactile lead. Error bars represent one standard error of the mean either side of the parameter values.

Figure 5.7 shows JND values averaged across observers. A second 2-way repeated measures ANOVA showed that JNDs did not differ significantly across sensory pairings ($F_{2, 6} = 0.93$, $p > 0.05$) and did not differ significantly between the three conditions (baseline and both polarities of asynchrony) ($F_{2, 6} = 2.37$, $p > 0.05$) within sensory pairings. In other words, JND did not differ significantly before and after asynchrony adaptation.

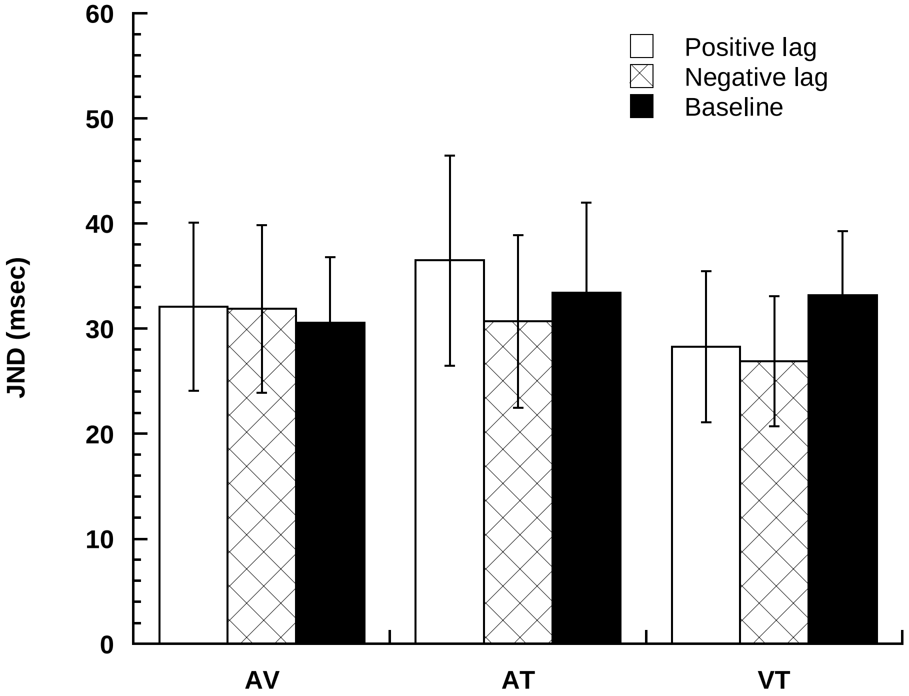


**Figure 5.7.** Average JND for each stimulus pairing, in baseline (solid filled bars), positive lag (unfilled bars) and negative lag (hatched bars) conditions. Positive and negative lags are defined as per Figures 5.3 - 5.5. Error bars represent one SEM either side of the parameter value.

### 5.2.5 Discussion

The results of the current study demonstrate that perceived timing is markedly influenced by recent experience, irrespective of the nature of the sensory pairing. Whilst the AV effects confirm previously published data (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Miyazaki et al., 2006; Vroomen et al., 2004), the AT and VT effects form the first convincing demonstration of adaptive temporal realignment in these modality pairings. These findings also provide evidence against the hypothesis that asynchrony adaptation effects are unlikely to be recorded in the tactile modality (Miyazaki et al., 2006). In addition, the JND data demonstrates that sensitivity to asynchrony is not affected by the process of adaptation to asynchronous stimulus pairs in *any* of the modality pairings or asynchrony polarities tested, a finding not previously demonstrated in the literature.

Although the data presented above clearly illustrate significant adaptive shifts in PSS following asynchrony adaptation in all three sensory pairings, a potential criticism is that observers could have unconsciously adopted a strategy of balancing their TOJs (e.g. 50% 'vision first', 50% 'sound first') during the method of constant stimuli, thereby forcing the PSS to be centred upon the mid-point of the range of temporal offsets between the stimuli. In other words, the effects demonstrated may be attributable to a combination of response bias and the offset range of SOAs used in the adapted conditions. Therefore, a control experiment was designed to eliminate this possibility.

In Section 3.2.3, a review of the features of various staircase paradigms was presented. As previously discussed, an advantage of the use of a procedure such as a PEST staircase (Taylor & Creelman, 1967) is that a *random* initial stimulus level (in this case, the SOA between the two stimuli) may be used. This is feasible because of the efficiency of the procedure in 'homing in' on the stimulus level around the value of interest (in this case, the PSS). If observers adopted a strategy of balancing their TOJs in such a case, the PSS value obtained with the staircase would be around the original *random* SOA value; in this case, the pattern of results presented above would not be replicated. On the other hand, if the results obtained with the use of a PEST staircase and initial SOA values chosen at random at the beginning of each run were in close correspondence with those presented above, the combination of response bias with an offset stimulus range in the TOJ task could be eliminated as a possible explanation for the results.

## 5.3 Perception of temporal order following adaptation to asynchronous stimulus pairs evaluated with a staircase paradigm

### 5.3.1 Methods and stimuli

Observers and stimuli used were the same as in the previous experiment (Sections 5.2.1 and 5.2.2). As before, observer CV was naïve as to the purpose of the experiment.

### 5.3.2 Procedures

The experimental methods were identical to the previous experiment (Section 5.2.3), with the exception that baseline and post-adaptation PSS was obtained via the use of a modified PEST staircase (Taylor & Creelman, 1967). The initial SOA was unknown to observers and varied randomly on a run-by-run basis between ±100msec. 50 test stimulus pairs were presented during each experimental run, with observers making a TOJ after each presentation. PSS was calculated as the mean of the presentation offsets after the first three reversals of the staircase. The adaptation and top-up experimental phases were identical to the previous experiment, with each test presentation offset being unknown to the observer and dependent on the staircase procedure. PSS values for each observer were averaged across three repetitions of the staircase procedure in each condition.

### 5.3.3 Results

Mean PSS shifts (relative to baseline) in the experiment are shown in Figure 5.8. A two-way repeated measures ANOVA revealed that, as before, the PSS shift was not significantly different across the sensory pairings ($F_{2, 6} = 3.7$, $p > 0.05$) but there was a highly significant effect of adaptation polarity ($F_{1, 3} = 100.7$, $p < 0.005$) without any significant interaction ($F_{2, 6} = 0.778$, $p > 0.05$). In other words, the pattern of PSS shifts showed a large degree of similarity to that in Section 5.2.4. The size of the recalibration effects measured with the PEST staircase paradigm were directly compared with the effects as measured using a method of constant stimuli paradigm with offset SOA ranges. This comparison revealed that there was no significant difference between the recalibration effects measured with the two paradigms ($F_{1, 3} = 0.796$, $p > 0.05$).

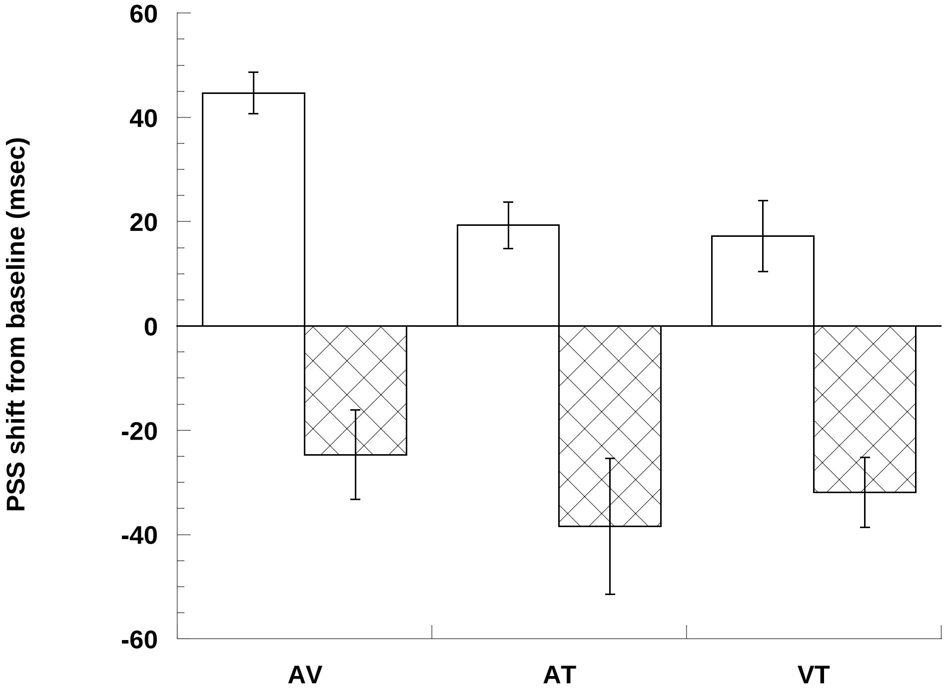


**Figure 5.8.** Average PSS shifts from baseline following adaptation to asynchronous stimulus pairs. All values (including baseline) were obtained by the use of a modified PEST staircase procedure. Unfilled bars relate to values obtained following adaptation to positive asynchrony, hatched bars following adaptation to negative asynchrony. Positive and negative lags are defined in an identical manner as Figure 5.6. Error bars represent one SEM either side of the parameter value.

### 5.3.4 Discussion

The results of the current experiment provide further evidence that exposure to asynchronous stimulus pairs causes shifts in perceived simultaneity in the direction of the adapting lag. The results also argue strongly against any potential concern that the effects demonstrated in Section 5.2.4 are caused by use of the method of constant stimuli associated with a strategy of response equalisation on the part of observers. If observers had simply balanced their proportion of responses from the start of the staircase, their final PSS values would faithfully reflect the original, *random* asynchrony value (between ±100msec) that initiated the staircase. Figure 5.8 shows this is definitively not the case. Thus, the presence of a robust recalibration of perceived time for all three sensory pairings is found to be consistent across different methodologies. The robustness of the observed effect is demonstrated by the close correspondence between the results of the current experiment and those presented in Section 5.2.4, with the recalibration effects not significantly differing between the two paradigms.

Subsequent to the present work, another study was published investigating asynchrony adaptation in the VT pairing with the use of a TOJ task (Keetels & Vroomen, 2008); the pattern of unchanged JNDs following asynchrony adaptation and PSS shifts in the direction of the adapting lag is essentially identical to that demonstrated in the present work.

Also subsequent to the present work, another study reported similar recalibration of VT PSS following exposure to asynchrony (Takahashi, Saiki, & Watanabe, 2008). In this case, a SJ (rather than TOJ) task was used, and the stimuli were computer-generated visual stimuli which appeared to be spatially aligned with a force-feedback device that provided tactile feedback. In this way, rather than using simple ‘flash/tap’ stimuli as in the present study and that of Keetels and Vroomen (2008), the experimenters were able to manipulate the temporal relationship between the application of force to a virtual haptic object, and the resultant apparent visual deformation of the object (Takahashi et al., 2008). As well as the shifts in perceived simultaneity following adaptation, asynchrony adaptation was found not to affect the estimate of observer sensitivity to asynchrony (Takahashi et al., 2008), as in the present work and that of Keetels and Vroomen (2008).

The fact that temporal recalibration (unaccompanied by any loss of observer sensitivity to asynchrony) in the VT pairing is observed using different paradigms (method of constant stimuli, staircase), judgment types (TOJ, SJ) and different stimuli (simple flash/tap pairs and naturalistic virtual objects) suggests that the observed shifts in perceived simultaneity are robust in nature. Therefore, research conducted independently of the present work (Keetels & Vroomen, 2008; Takahashi et al., 2008) provides strong support for the findings presented here.

In contrast to the present work, a previous study by Navarra et al. (2007) found no significant difference between AT PSS measured after adaptation to synchronous and asynchronous (‘sound leads touch’) stimulus pairs. Rather, they found that asynchrony adaptation induced a small (12msec) increase in JND relative to their ‘adapt synchronous’ condition. This is the exact opposite finding to that presented here and therefore requires consideration. Numerous methodological differences between the two studies make comparison problematic. For example, the magnitudes of adaptive PSS shifts in the present study are judged relative to an unadapted baseline whereas Navarra et al. (2007) used a physically synchronous AT pairing as their baseline against which to compare their adapted results. Given the documented dissociation between physical and perceived synchrony in the AT domain (Navarra et al., 2007; Zampini et al., 2005a) a *physically* synchronous audiotactile adapting stimulus may or may not be *perceptually* synchronous (e.g., Figure 5.4 shows observer JVMH to have an unadapted AT PSS of -18msec). The use of adaptation to physically synchronous stimulus pairs in the ‘baseline’ condition may also have affected the ‘baseline’ PSS values used as a comparison by Navarra et al. (2007) – previous work in the AV pairing indicates that PSS measured without adaptation differs significantly from that measured following adaptation to physically synchronous stimulus pairs (Fujisaki et al., 2004). If this finding were to also apply to the AT pairing it is possible that, had unadapted PSS been measured and compared with the adapted ‘sound leads touch’ data, statistically significant temporal recalibration effects would have been measured by Navarra et al. (2007).

Another significant difference is Navarra et al.’s alteration of stimulus characteristics between adaptation and test phase; the adapting auditory and tactile stimuli were 250Hz tones/vibrations, whereas test stimuli were 300Hz. If adaptation modulates perceived timing at a relatively early processing stage, the manipulation of such low-level stimulus characteristics may prevent transfer of PSS shifts between the two experimental phases. However, caution must be applied when making assumptions about the neural locus of asynchrony adaptation on the basis of such evidence: the data presented here suggest a relatively late-stage mechanism that recalibrates *any* sensory input with respect to recent experience. Specifically, the size of the effects in the present work is broadly comparable across modality pairings and asynchrony polarity, suggesting that the recalibration mechanism may modulate the same fixed proportion of the adaptation asynchrony (mean 36.2%, SEM 2.6% when pooled across Method of Constant Stimuli and PEST methodologies) regardless of the modality of the signal itself. Whether the recalibration effects remain a constant proportion of adapting SOA when the adapting SOA is altered remains to be elucidated in detail, although it is clear that modulating the adapting asynchrony affects the magnitude of temporal recalibration in the AV domain (Fujisaki et al., 2004).

It should also be noted that Navarra et al. (2007) used adapting stimulus pairs with 75msec asynchrony, smaller than that used in the present work and other studies (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Keetels & Vroomen, 2008; Vroomen et al., 2004). At the present time, the minimum amount of asynchrony required for adaptive PSS shifts is undocumented; it therefore remains possible that 75msec of adapting asynchrony is insufficient to induce *statistically significant* temporal recalibration in the AT pairing. Although it is not possible to speculate with confidence which (if any) of these factors are responsible for the lack of temporal recalibration recorded by Navarra and colleagues, clear methodological differences with the present work are obvious and a direct comparison of the two studies is therefore problematic.

Another study published subsequent to the present work systematically investigated the effects of exposure to asynchronous AV, AT and VT stimulus pairs on perceived simultaneity in the AV, AT and VT pairings; in total, nine combinations of adapt and test modality combinations was used (Harrar & Harris, 2008). This study found that adaptation to AV, AT or VT asynchronous stimulus pairs did not significantly affect PSS (measured using a TOJ task) in the AT or VT pairings. In contrast, adaptation to AV, AT or VT asynchronous stimulus pairs shifted AV PSS relative to an unadapted baseline. JND was unaffected by asynchrony adaptation, except in the adapt AT test AV combination where JND was reduced following adaptation (Harrar & Harris, 2008). The study concluded that the AV asynchrony adaptation system is fundamentally different to the AT and VT systems in that it allows adaptive PSS shifts. Some of the findings of this study agree with the present work (adaptive AV PSS shifts with unchanged JND following adaptation to asynchronous AV stimulus pairs; unchanged AT and VT JND following adaptation to asynchronous AT and VT stimulus pairs, respectively). Conversely, the finding (Harrar & Harris, 2008) that AT and VT PSS is unaffected by exposure to AT and VT asynchronous stimulus pairs, respectively, is in disagreement with the present work. Clearly, the finding that the AV asynchrony adaptation mechanism is qualitatively different to the AT and VT mechanism(s) also differs with the conclusions presented in the current study, which show no significant differences between the three modality pairings.

As with the work of Navarra et al. (2007), the precise reason(s) for this difference between the present study and that of Harrar and Harris (2008) are currently unclear. However, one factor likely to be significant is the use by Harrar and Harris of only a single adaptation polarity for each of their modality pairings (e.g., adapting to a physical lead of sound over touch); this potential confound also being present in the work of Navarra et al. (2008). Clearly, a more thorough approach would be to test *both* polarities of asynchrony in each condition – in such a situation, the authors could simply have compared PSS values obtained following both polarities of asynchrony adaptation to check for significant differences. Although comparison of a single adapted condition with its unadapted counterpart is undoubtedly instructive, the comparison between PSS values obtained following adaptation to opposite polarities of asynchrony is necessary to definitively reject the existence of asynchrony adaptation effects[8]. Two previous studies were able to confirm such effects in the AV domain without presenting any baseline PSS values whatsoever (Keetels & Vroomen, 2007; Miyazaki et al., 2006), although these represent isolated examples within the literature. Because of this potential concern and the results of the present study and others (Keetels & Vroomen, 2008; Takahashi et al., 2008), the conclusion of Harrar and Harris (2008) that only AV PSS can be shifted by asynchrony adaptation should be treated with caution.

However, the fundamental question addressed by Harrar and Harris - whether asynchrony adaptation influences subsequent TOJs in non-adapted stimulus pairings – is undoubtedly worthy of investigation. The results of such a study should provide further clues as to the nature of the asynchrony adaptation mechanism(s). At least three potential outcomes of such an experiment are possible, summarised in Figure 5.9(a-c) using the example of the AV pairing (adaptation to a physical lead of sound over vision).

---

[8] Note that in the present work, no such comparison was necessary because of the statistically significant difference in PSS values obtained following both polarities of

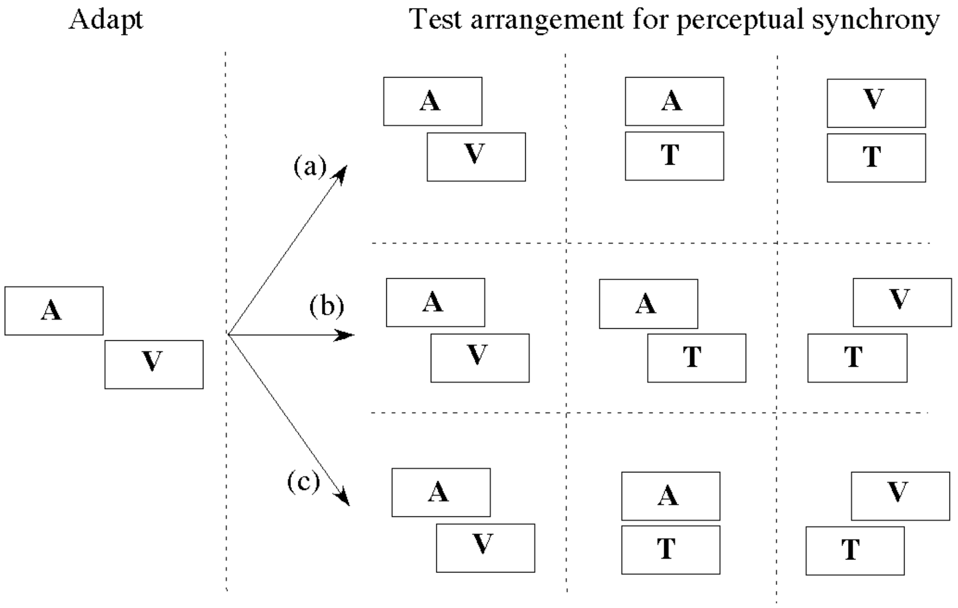


**Figure 5.9(a-c).** Possible outcomes when adapting to asynchronous stimulus pairs presented in one modality (in this example, AV (sound leads vision)) and measuring PSS in the other sensory pairings. For simplicity, baseline PSS is assumed to be zero (representing a veridical perception of AV temporal order). Adaptation to a lead of sound over vision causes observers to require a physical lead of sound over vision for perceptual simultaneity (a, b and c, left column). At least three different outcomes are possible when measuring PSS in the AT and VT pairings (a, b and c, middle and right column respectively), which are summarised in the text.

Figure 5.9(a-c) schematises the possible effects on AV, AT and VT PSS following adaptation to AV asynchronous stimulus pairs (sound leading vision). The first possibility, shown in the top panel (a), is that AV PSS is altered by adaptation such that a lead of sound over vision is now required for perceptual synchrony, but AT and VT PSS is unaffected. This would suggest a temporal recalibration mechanism specific to the adapting stimulus pairing.

---

asynchrony adaptation compared to baseline (unadapted) PSS.

A second possibility is shown in the middle panel (b): the nervous system may treat adaptation to a physical lead of (e.g.) sound over vision as equivalent to a physical *lead* of sound relative to any other modality, and a physical *lag* of vision relative to any other modality. In such an instance, we would expect observers to require a physical lead of sound over touch for perceived AT simultaneity, and a lead of touch over vision in the VT domain. This outcome would be consistent with a single supramodal asynchrony adaptation mechanism operating at a relatively high level of the perceptual system (most likely beyond primary sensory cortices), as suggested by the present work.

Although the PSS shifts following adaptation described here appear robust, one question arising from the present study is whether both sensory modalities in any pairing undergo shifts in perceived time of approximately equal magnitude, or whether the degree of temporal recalibration is asymmetric across modalities. Changing the stimulus pairing between adaptation and test phases (as in Figure 5.9) may enable this detail to be elucidated. The possibility that *both* modalities recalibrate by a similar amount can be seen in the middle panel (b).

A further possibility is shown in the lower panel (c). Consider the example of adaptation to a lead of sound over vision illustrated in Figure 5.9. Given the relative temporal reliability of auditory information (Sections 2.1.2, 2.1.4 and 2.1.5), it is possible that when testing the AT pairing minimal temporal recalibration is evident – in other words, the auditory modality is unaffected by asynchrony adaptation. When testing the VT pairing, adaptation may shift

PSS such that a physical lead of touch over vision is required for perceived simultaneity due to the relatively poor temporal reliability of visual processing (Sections 2.1.2, 2.1.4 and 2.1.5). A recent conference presentation suggested that in fact the *auditory* stimuli are shifted in perceived time: adaptation to AV asynchrony modulated AV and AT, but not VT, PSS (Di Luca, Machulla, & Ernst, 2007). Such a finding could reflect the effect of observer-source distance on AV perceived timing: because of the speed of light, visual information reaches an observer effectively instantaneously irrespective of distance, whereas sound is increasingly delayed in arrival at an observer with increasing distance. Thus, it may be that it is simply unnecessary for visual signals to be temporally recalibrated, in contrast to auditory signals which would presumably benefit from a degree of temporal flexibility. Whether auditory or visual sensory input is temporally recalibrated, previous work in the temporal (Roach et al., 2006) and spatiotemporal (Heron et al., 2004) AV domains raises the strong possibility that manipulating the relative temporal reliabilities of stimuli (via manipulation of parameters such as onset profile) may enable *both* modalities to temporally recalibrate as their reliability becomes more equivalent.

In the present study, JNDs were found to be invariant across modality pairings and adapted vs. baseline conditions. As such, it appears the nervous system has the ability to adaptively recalibrate sensory temporal relationships without a discernable loss of sensitivity. Anecdotally, observers reported that by the end of the adaptation phase the physically asynchronous stimulus pairs felt markedly closer to being perceptually synchronous. The JND data suggests this phenomenon is not a product of a progressive loss in sensitivity – rather, the signals (relative to one another) are subject to an adaptive temporal recalibration. It is therefore unlikely that an asynchrony detection circuit such as that described previously in the AV domain (Bushara et al., 2001) plays a significant role in the effects described here. If this were this the case, it is more likely that exposure to repeated asynchronous AV stimulus pairs (as in the adaptation phase) would fatigue the asynchrony detection circuit, increasing the difficulty of subsequent relative timing judgments and consequently increasing JND.

However, the perceptual recalibration observed in the present work is consistent with Helson's adaptation-level theory (Helson, 1964). In Helson's framework, human sensations are judged relative to an aggregate of recent experience which is constantly updated by novel sensory stimulation (Behar & Bevan, 1961). For example, adaptation-level theory predicts that an object of any given weight is perceived as heavy if a subject has recent experience of carrying objects that are relatively light, and vice versa. In other words, objects of the same objective weight may be perceived as heavier or lighter depending on recent experience, which acts as a perceptual 'anchor' against which new sensation is judged. In the case of the present work, it seems likely that the fixed adapting asynchrony provides observers with a revised perceptual 'anchor' (e.g., sound leads vision) around which to centre their adaptation level.

The magnitude of the PSS shifts displayed in the present work as a proportion of the adapting lag (mean 36.2%) is greater than in the majority of other studies, which found adaptive PSS shifts ranging between 3.5-12.9% of the adapting asynchrony (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Keetels & Vroomen, 2008; Takahashi et al., 2008; Vroomen et al., 2004). Similar to the present work, however, Harrar and Harris (2008) found average AV PSS shifts of 32% of their adapting asynchrony (100msec). Similarly, another study found that observers adapting to a lead of vision over sound underwent PSS shifts of approximately 50% of the adapting asynchrony (Heron et al., 2007). The reason for these discrepancies is investigated in Section 6.5.

Returning to the hypothesis of Miyazaki et al. (2006), the adaptive shifts presented here argue against the idea that the effects observed here and by others (Fujisaki et al., 2004; Heron et al., 2007; Vroomen et al., 2004) are peculiar to the AV domain. Miyazaki et al. (2006) speculate that if the perceptual consequences of repeated exposure to tactile signals are to be explained within a Bayesian framework, subjects will impose perceptual asynchrony onto physically simultaneous signals. This asynchrony will *not* be a classic negative aftereffect such as that observed here (e.g., following adaptation to 'vision leads sound' stimulus pairs, observers are more likely to respond 'sound first' when subsequently presented with physically synchronous stimulus pairs, as seen in Figure 5.1); rather, it will be similar to that of the adapting stimulus. In other words, exposure to 'vision leads sound' pairs would cause observers to respond 'vision first' to subsequent physically synchronous stimulus pairs: consequently, the psychometric functions would be laterally displaced in the *opposite* direction to that shown in Figure 5.1. Both Bayesian[9] adaptation and asynchrony adaptation may be active, competing mechanisms, with the perceptual outcome dependent on which modalities are tested (Miyazaki et al., 2006). In the case of AV asynchrony, the asynchrony adaptation mechanism has been proposed to overcome the magnitude of any Bayesian recalibration. Alternatively, when tactile signals are involved Bayesian-type recalibration dominates. The latter outcome has been attributed to observers adopting a new prior assumption of tactile temporal order. This prior corresponds to the asynchrony to which observers are repeatedly exposed to during the adaptation phase (Miyazaki et al., 2006). This concept has received support from a recent VT study showing that new priors can be adopted by observers given just a single hour of training (Ernst, 2007).

Alternatively, it is possible that the results of both studies may be compatible with Bayesian models of perception. Given that approximately simultaneous taps to one hand followed by the other (as employed by Miyazaki et al (2006)) are a relatively infrequent event in a natural environment, observers are unlikely to possess deeply ingrained prior assumptions about their temporal order. In this situation, observers may be more likely to modify any pre-existing prior assumptions about the temporal relationship between the signals (Ernst, 2007). In contrast, it is reasonable to speculate that a lifetime's experience of close temporal correlation may produce a strong prior assumption of synchrony. As discussed in Section 5.1.1, this would be the

---

[9] Although the description 'Bayesian adaptation' is ambiguous in this context, the terminology is employed here as it was originally used by Miyazaki et al. (2006).

case for both proximal AV events and *all* AT and VT events whose sensory components must – by definition – occur simultaneously and arrive at their respective receptor surfaces approximately simultaneously. Post-adaptation 'rebound' effects of the type described here could well be the product of a tendency to realign perception in a manner that maintains concordance with prior assumptions. A recent study has proposed that adaptation may alter the likelihood function/perceptual outcome without influencing any pre-existing prior assumptions (Stocker & Simoncelli, 2006). This implies that when observers *expect* synchronous sensory signals they realign their perceived sensory timing rather than their pre-existing prior knowledge of the environment.

In conclusion, the results of the current study demonstrate a degree of temporal flexibility that appears comparable across the range of modality pairings investigated. This cannot be explained simply in terms of maintaining veridicality in the time domain. Rather, it seems that the nervous system may realign *any* temporal signal that fails to meet its prior assumptions about the world. Of course, the task of formulating a sensory estimate of 'when' is far from abstract - our motor commands are only as accurate as the sensory estimates that guide them. A critical question for future work is how purely sensory adaptation (as described in the current study) relates to recent examples of sensorimotor temporal recalibration (Pesavento & Schlag, 2006; Stetson et al., 2006).

## 5.4 Analysis of observer sensitivity and transitivity of temporal order judgments

### 5.4.1 Introduction

Previously, analysis of TOJ data showed that observer sensitivity to asynchrony is approximately the same in the AV, AT and VT pairings (Section 4.3.4), confirming the suggestion made by earlier researchers (Hirsh & Sherrick, 1961). Given the well-documented differences in temporal processing within the three modalities (Sections 2.1.1-2.1.5) this may be considered surprising; however, it supports the notion that crossmodal TOJs are mediated by a single decisional mechanism serving the three modality pairings.

If TOJs are mediated by a common neural mechanism, the possibility arises that crossmodal PSS may be *transitive*; that is, for any given observer, PSS values in any two of the modality pairings should predict PSS in the third pairing. For example, consider an observer who requires a physical lead of vision over sound of 15msec for perceptual simultaneity, and a lead of sound over touch of 10msec. If crossmodal PSS is transitive, then VT PSS should correspond to a 25msec lead of vision over touch. This outcome would suggest not only that crossmodal TOJ was mediated by a single decisional mechanism, but also that perceived event timing for any single modality is unaffected by the processing of stimuli in another modality. Conversely, should PSS be intransitive, this could be interpreted as reflecting the fact that crossmodal TOJ processing is accomplished by multiple independent mechanisms, and/or that latency for a given modality is dependent on which modality it is paired with.

Thus far, the transitivity issue has received surprisingly little attention. Recent work by Machulla and colleagues showed that crossmodal PSS was intransitive, consistent with separate neural mechanisms and/or latency of stimuli in one modality being dependent on which modality it is paired with (Machulla et al., 2007). However, within the visual domain, judgments of simultaneity between three stimuli (identified by their different positions in the visual field) were found to be transitive (Corwin & Boynton, 1968). These authors interpreted their result as being indicative of a visual ‘simultaneity centre’ within the brain. This conclusion is hard to apply to the present work due to the unimodal visual paradigm employed and the use of a SJ (rather than TOJ) task by Corwin and Boynton (1968). Additionally, measures of observer sensitivity derived from SJ data may reflect observers’ response criterion as well as sensitivity to asynchrony (Section 3.3.3). Analysis of PSS and TOJ JND values may thus help to establish whether or not multimodal TOJ is mediated by a single mechanism, and the degree of inter- or independence between latencies in the V, A and T modalities. With this goal, all TOJ data obtained thus far were pooled and reanalysed, along with additional data sets from two more naïve observers.

### 5.4.2 Methods & stimuli

The TOJ data obtained from author JVMH, DW, JH and DS (Section 4.3.4), and AA and CV (Section 4.5.4), was combined with new data sets obtained by naïve observers CD and DL using the methods, stimuli and procedures described previously (Section 4.3.3). Prior to data collection, CD and DL undertook approximately three hours TOJ practice (spread over one week) to facilitate learning of the TOJ task; this data was excluded from analysis.

Predicted VT PSS values were calculated for each observer from AV and AT PSS values. For example, for author JVMH AV PSS corresponded to a 10.6msec physical lead of sound over vision, and AT PSS to a 17.7msec lead of touch over sound. Therefore, if crossmodal PSS is transitive, VT PSS for this observer should correspond to a 28.3msec lead of touch over vision (-28.3msec in the convention employed in the present work). The decision to predict VT PSS (rather than AV or AT) was entirely arbitrary, but this is inconsequential as if crossmodal PSS is transitive then *any* two PSS values should be able to predict the remaining PSS value.

### 5.4.3 Results

PSS values for each observer are shown as Figures 5.10 - 5.12 (corresponding to the AV, AT and VT pairings, respectively).

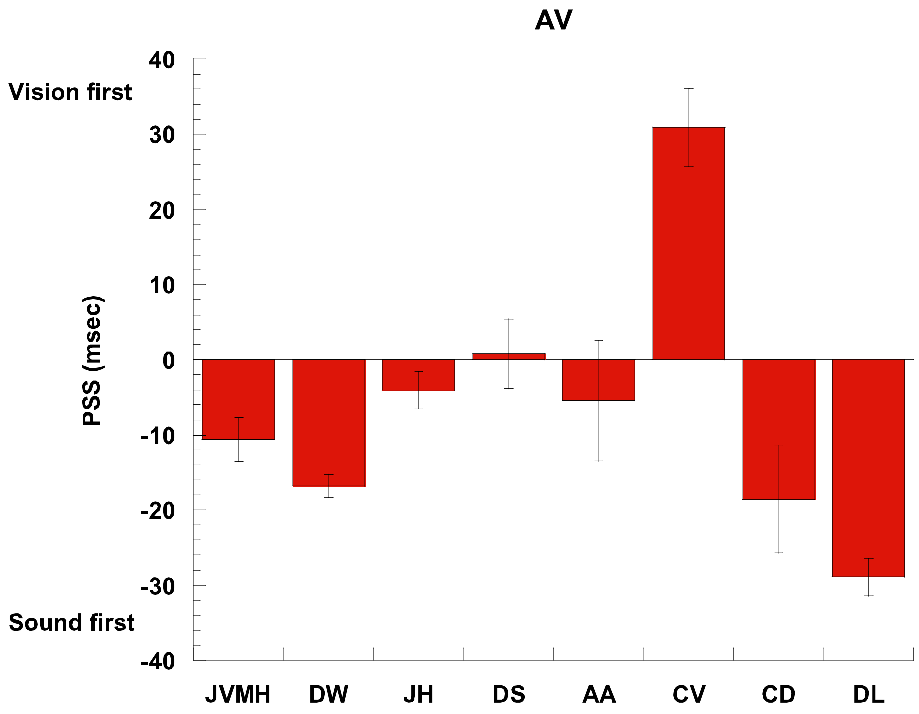


**Figure 5.10.** AV PSS for 8 observers. Positive values correspond to a physical lead of vision over sound. Error bars represent one standard deviation either side of the parameter values.

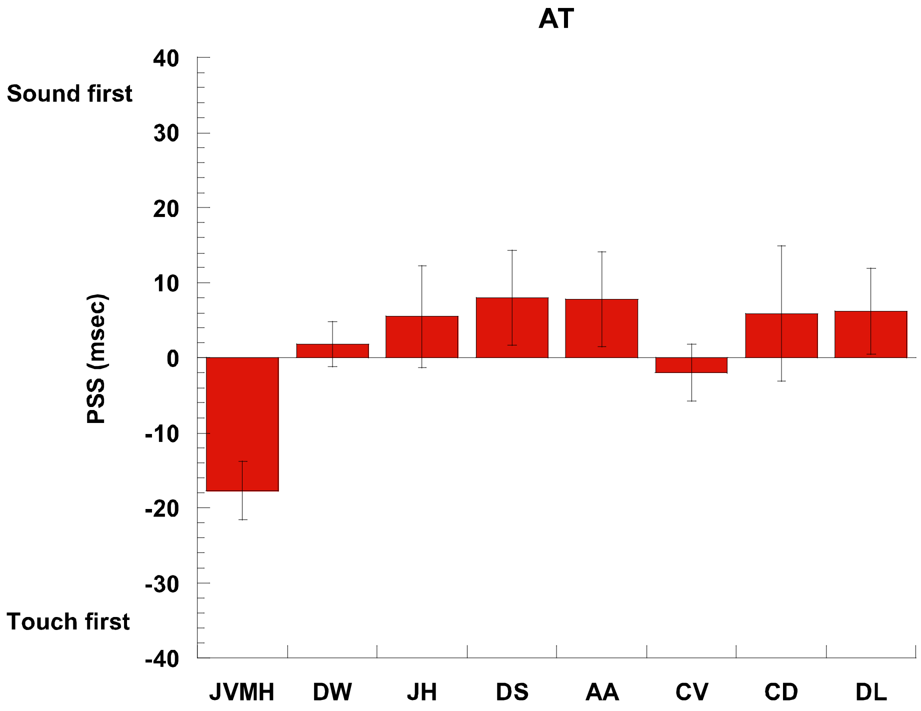


**Figure 5.11.** AT PSS for 8 observers. Positive values correspond to a physical lead of sound over touch. Error bars represent one standard deviation either side of the parameter values.

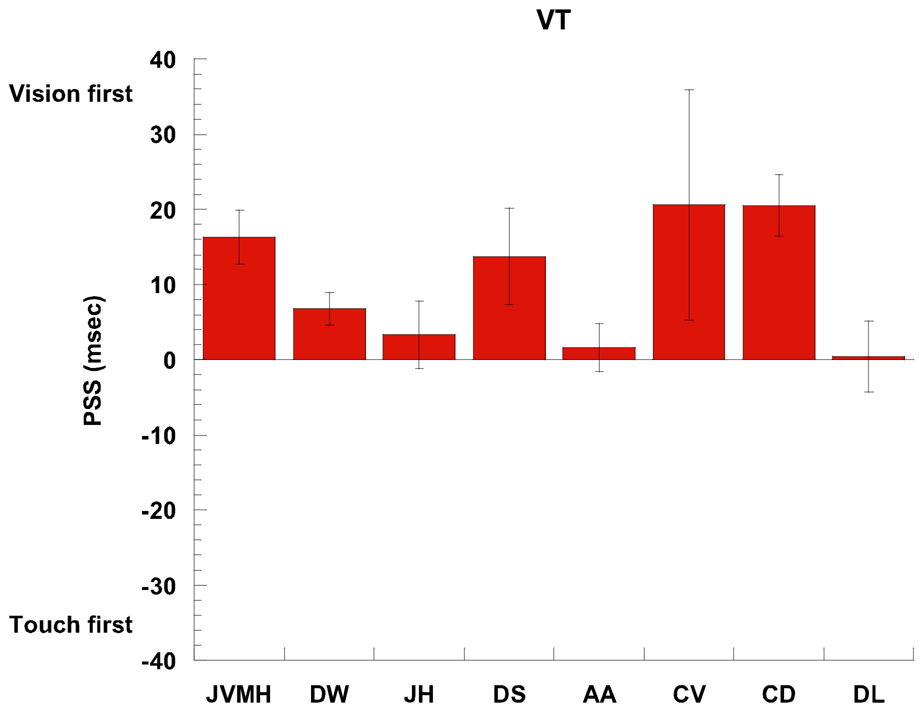


**Figure 5.12.** VT PSS for 8 observers. Positive values correspond to a physical lead of vision over touch. Error bars represent one standard deviation either side of the parameter values.

In Figure 5.13, VT PSS values predicted from AV and AT values by assuming transitivity of crossmodal PSS are plotted against VT PSS as measured with a TOJ task for each observer. If crossmodal PSS is transitive, data points should lie on the diagonal line (gradient = 1). Four data points of eight meet this criterion at the 95% confidence limits. Comparison of predicted (transitivity assumed) and measured VT PSS showed a linear correlation coefficient of 0.15 ($p = 0.72$).

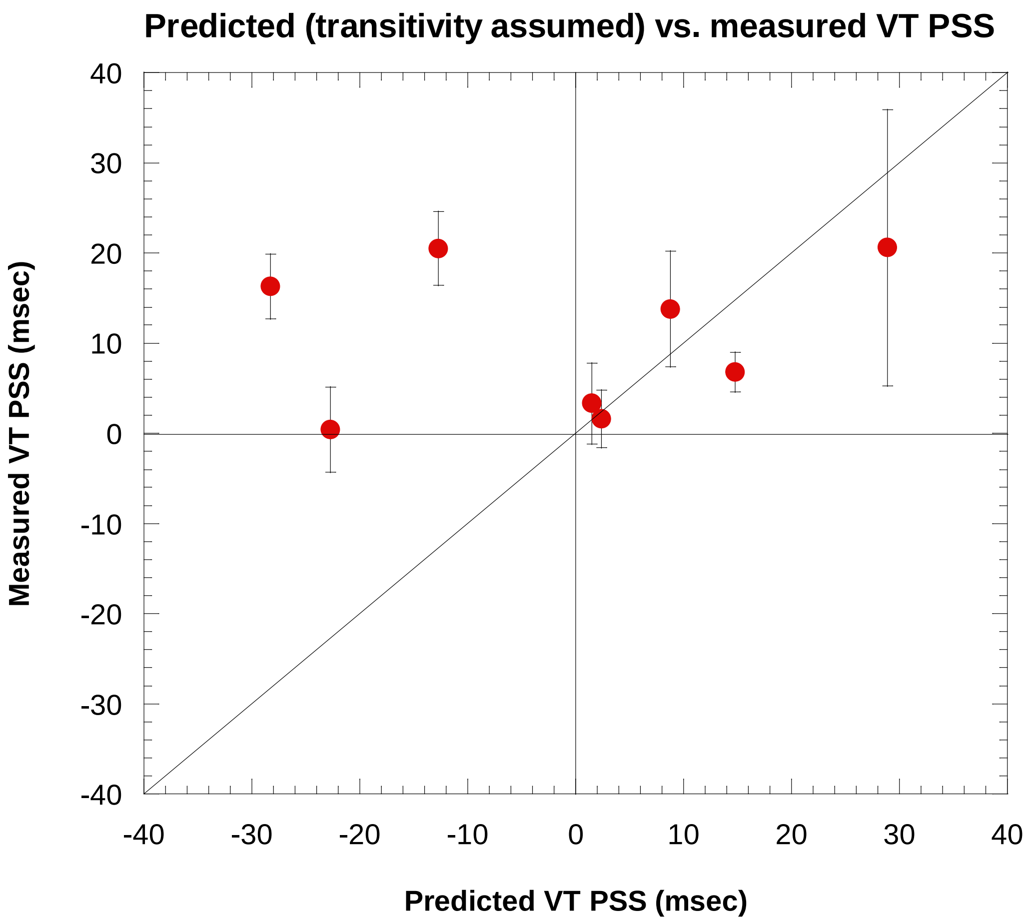


**Figure 5.13.** VT PSS predicted by assuming transitivity of crossmodal PSS compared with that measured using a TOJ task. Should crossmodal PSS be transitive, data points should lie on the diagonal (gradient = 1). Error bars represent one standard deviation either side of the parameter values.

A one-way repeated measures ANOVA on the JND data showed that JND values did not differ significantly between the modality pairings ($F_{2, 14} = 1.44$, $p > 0.05$). In other words, for each observer JND was approximately the same in the AV, AT and VT conditions.

### 5.4.4 Discussion

Figures 5.10 - 5.12 generally support the hypothesis that PSS is observer-specific, previously suggested in the AV domain (Stone et al., 2001). However, inter-observer differences are particularly obvious in the AV pairing (Figure 5.10), with PSS ranging from a visual lead of 31msec (observer CV) to an auditory lead of 29msec (observer DL). As summarised previously (Section 2.1.3), the available literature on this topic is similarly dichotomous.

In the AT pairing, conversely, seven observers from eight display PSS values that do not differ significantly from zero at the 95% level (i.e., PSS values are within 1.96SD of zero). In this respect, the results presented here differ slightly from previous studies which have been unanimous in finding that a lead of touch over sound is required for perceived simultaneity (Dinnerstein & Zlotogura, 1968; Harrar & Harris, 2008; Hirsh & Sherrick, 1961; Machulla et al., 2007; Navarra et al., 2007; Zampini et al., 2005a).

All eight observers in the VT condition require a physical lead of vision over touch for perceived simultaneity, as would be expected from transduction and neural latencies (Section 1.1.1 and 1.1.3) and simple RT values (Sections 2.1.1 and 4.2.4). However, note that this visual lead does not differ significantly from zero for five of the eight participants. As in the AV pairing, previous studies have produced conflicting results regarding PSS (Section 2.1.3); the reason(s) for this remains unknown.

It is difficult to draw any firm conclusions regarding transitivity of crossmodal PSS from the data shown in Figure 5.13; four observers of eight show results consistent with transitivity, whereas the remaining four observers do not. However, given the results of the correlation analysis and the previous work of Machulla et al. (2007), the more likely outcome is that PSS is intransitive. Machulla and colleagues, using 16 observers, suggested that their results indicated that either processing time in one modality was dependent on the modality it was paired with in the crossmodal TOJ task, or that there is no common (supramodal) TOJ mechanism within the brain. However, the analysis of JND in the present work, which shows that sensitivity to asynchrony does not differ between the modality pairings, is consistent with such a supramodal mechanism, as is the previous data of Hirsh and Sherrick (1961). A caveat remains that the relatively small sample of observers in the present study may cause ANOVA to have insufficient statistical power to find genuine differences in JND between the sensory pairings.

The apparent intransitivity of crossmodal PSS combined with statistically indistinguishable JND data in the three sensory pairings (but see caveat above) can be interpreted in one of at least two ways. Firstly, at least three distinct mechanisms (AV, AT and VT) for perception of temporal order exist within the nervous system, which have approximately equal sensitivity to temporal order. This hypothesis is unlikely as it does not take into account the different temporal processing capabilities of the sensory modalities (Section 2.1.2). Specifically, if this hypothesis were correct then a more intuitive pattern of results would feature the AT pairing having higher sensitivity to asynchrony than the pairings featuring the visual modality. Such a pattern of results has been previously demonstrated using SJ (Fujisaki & Nishida, 2007a) and asynchrony matching (Sinex, 1978) tasks, but not with a TOJ task.

The second possibility is that there is a common mechanism for processing temporal order in the AV, AT and VT pairings within the nervous system, but that the processing time for each modality is dependent on the paired modality; in other words, temporal order processing within the sensory modalities is not performed independently. Such interdependence of temporal processing between the modalities has previously been suggested in the AV domain on the basis of observers' duration judgments (van Wassenhove, Buonomano, Shimojo, & Shams, 2008), but not with crossmodal TOJs. However, given the numerous neurophysiological interactions between the senses at all levels of the central nervous system reviewed elsewhere (Calvert, 2001; Schroeder et al., 2003; Shimojo & Shams, 2001; Stein & Meredith, 1993; Stein & Stanford, 2008), such links are certainly credible.

In conclusion, the overall weight of evidence is insufficient to support the hypothesis that crossmodal PSS is transitive. Conversely, the hypothesis that crossmodal TOJ is mediated by a single neural mechanism is credible given the JND data. Thus, it is likely that the latency of sensory events (in a crossmodal context) is dependent on the modality which is paired with it. Further work is needed to elucidate the precise nature and neural locus of these crossmodal interactions. Additionally, neuroimaging studies would be invaluable in confirming or rejecting the hypothesis that a single neural mechanism mediates crossmodal TOJ.

# Chapter 6

## 6.1 Adaptation to spatially disparate asynchronous audiovisual stimulus pairs

### 6.1.1 Introduction

It is clear from the data presented previously within this thesis (Sections 5.2.4 and 5.3.3) and elsewhere (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Vroomen et al., 2004) that perception of audiovisual (AV) temporal order in humans may be modified by a period of adaptation. Specifically, adaptation to asynchronous AV stimulus pairs (vision physically leading sound) causes the Point of Subjective Simultaneity (PSS) to shift such that a physical lead of vision over sound is now required for perceptual simultaneity (Figure 5.1). Adaptation to a physical lead of sound over vision causes PSS to shift in the opposite direction, such that a physical lead of sound over vision is required for perceived simultaneity.

Fujisaki et al. (2004) have presented data showing that this temporal recalibration is unaffected by using different stimuli and/or tasks in adapt and test phases. The observed effects are also unaffected by changing the laterality of auditory stimuli between adapt and test phases (Fujisaki et al., 2004). Given these findings, it seems reasonable to assume that the AV asynchrony adaptation mechanism operates at a relatively late stage of the neural processing hierarchy. If the mechanism operated at an early stage of the perceptual system, it is much more likely that changing such low-level stimulus features (e.g., auditory stimulus laterality) between adapt and test phases would result in minimal or absent temporal recalibration. Evidence presented elsewhere in this thesis (Sections 5.2.4 and 5.3.3) showing equivalent recalibration effects and JND in the AV, audiotactile (AT) and visuotactile (VT) modality pairings, is also suggestive of a single late-stage mechanism for recalibration of perceived temporal order following exposure to asynchronous stimulus pairs.

Currently, the spatial specificity of asynchrony adaptation effects is unclear. One possibility is that the effects are unitary and generalise across space. Such a lack of spatial specificity is consistent with a single, late-stage recalibration mechanism such as that discussed above. If this were the case, adapting to a lead of (e.g.) vision over sound at one location would be expected to recalibrate TOJs made at *any* location such that PSS following adaptation corresponded to a physical lead of vision over sound. Additionally, were observers to adapt to two *opposing* polarities of AV asynchrony (e.g., ‘vision leads sound’ at one location, ‘sound leads vision’ at another location), it is likely that the opposing polarities of the adapting asynchrony would cancel each other out, resulting in minimal or absent temporal recalibration at both locations.

However, other temporal aftereffects appear to be confined to the adapted region of external space. For example, compressions of perceived visual duration following adaptation have been shown to occur in a highly spatially-specific manner (Johnston et al., 2006): Johnston and colleagues showed that adaptation to 20Hz oscillatory motion or visual flicker reduced the perceived duration of subsequent visual stimuli presented at the same location (relative to unadapted locations). Although it has been claimed that these effects are spatiotopic in nature (Burr, Tozzi, & Morrone, 2007), subsequent work has suggested that the effects are in fact retinotopic (Bruno, Ayhan, & Johnston, 2008). Similarly spatially-specific temporal distortions of perceived duration following adaptation have also been recorded in the tactile domain: adaptation to 35Hz vibration delivered to a finger reduced perceived stimulus duration at the adapted finger, but not at an unadapted finger (Watanabe et al., 2008). The spatial specificity of these temporal aftereffects is more consistent with early-stage temporal recalibration mechanisms.

If the spatial specificity of the effects outlined above also applies to the asynchrony adaptation mechanism, it is likely that adaptation to asynchronous AV stimulus pairs will not influence TOJs measured at an unadapted location. Such spatial specificity may also allow adaptation to opposing polarities of AV asynchrony at different spatial locations to produce opposing PSS shifts at these locations. In other words, concurrent adaptation to ‘vision leads sound’ and ‘sound leads vision’ stimulus pairs at different spatial locations may cause concurrent PSS shifts in the direction of the adapting lag at each location.

With these factors in mind, the proposed experiment aimed to investigate whether observers are able to recalibrate perceived AV temporal order at two different locations, and whether this recalibration may occur in opposing directions concurrently. To accomplish this, it was necessary to use different experimental apparatus to that used previously (Chapters 4 and 5), as the auditory and visual stimuli of each stimulus pair were required to be spatially co-localised. These requirements were met through the use of individually-measured Head-Related Transfer Functions (HRTFs) for each observer. Further discussion of the motivation for using HRTFs and a brief description of the underlying theory behind such stimuli is presented below.

## 6.2 The Head-Related Transfer Function (HRTF)

Historically, AV experiments have typically employed one of at least two types of visual stimuli: a simple flash of light generated by a Light Emitting Diode (LED) or some other light source, or stimuli with more complex spatial and/or temporal characteristics (e.g., a Gabor patch). Such complex stimuli are usually computer-generated and presented on a monitor. If the experimenter wishes to also use auditory stimuli within the study, the situation becomes more complicated. One option is to simply present the auditory stimuli over headphones or a loudspeaker. This is acceptable if the experiment does not require spatial co-localisation of auditory and visual stimuli – for example, an AV TOJ task could involve an observer judging which of a tone presented over headphones or an on-screen visual stimulus came first (as in Section 4.3.3).

However, if the experimenter is interested in whether observers can recalibrate perceived AV timing at different locations, it is vital that the auditory and visual stimulus pairs are both able to be presented at perceptually compelling spatial locations. In this case, presentation of standard auditory stimuli over headphones is not possible because such stimuli are typically perceived as originating inside the listeners' head (Wightman & Kistler, 1989a). An alternative is to present auditory stimuli over separate loudspeakers, whilst presenting visual stimuli over LEDs aligned with the speakers. However, the use of external speakers in these circumstances precludes the use of computer-generated visual stimuli presented on a monitor (as the loudspeakers cannot be co-localized in space with a region of the computer monitor screen). Thus, any experiments using such experimental apparatus would be limited to the use of LEDs (or similar sources) as visual stimuli. An additional disadvantage of such apparatus is the relative lack of portability of external loudspeakers and amplification in comparison with headphones.

Another approach is to use headphone presentations of auditory stimuli which *appear* to originate from the same location as the visual stimuli. This feat may be accomplished through the use of auditory stimuli containing spatial cues obtained by measurement of an individuals' HRTF. A HRTF describes how a sound source of a given spectral content and spatial location is filtered by the pinnae, head and torso of an observer. It is known that the primary auditory spatial cues utilised by humans are differences in the arrival times of sounds arriving at the left and right ear canal (interaural time differences (ITD)), differences in the intensity of sounds arriving at the left and right ear canal (interaural intensity differences (IID)), and the spectral content of sounds in the two ear canals after filtering by the external ears, head and torso (Blauert, 1997). It is also possible to conceptualise ITDs as Interaural Phase Differences (IPDs), as soundwaves arriving *in phase* at each ear from a lateralised source will do so at different times (i.e., with an ITD), whereas sound waves may also arrive at the same time at each ear but out of phase (i.e., with an IPD); for the sake of simplicity, ITDs will be used throughout this discussion.

With low frequency sound (with a frequency less than approximately 1000Hz), ITDs are the dominant cues used for localisation, whereas with high frequency sound IIDs predominate (Middlebrooks & Green, 1991; Rayleigh, 1907). This can be seen in Figure 6.1; only at sufficiently high sound frequencies (i.e., short wavelengths) will the head create an ‘acoustic shadow’ and thus an IID. The magnitude of this shadowing effect is dependent on head size and the wavelength of the incident sound waves (Middlebrooks & Green, 1991).

By measuring a sound of known laterality and spectral content as it arrives at each ear - using microphones placed in the ear canals - the precise arrival time, intensity and spectral content of the incident sound at each ear can be calculated to give the HRTF for that particular incident angle of sound (Wightman & Kistler, 1989a). A given stimulus convolved with the HRTF corresponding to a given location and presented over headphones should therefore contain all the spatial cues available to an observer during natural, free-field listening. Thus, such a stimulus/HRTF combination will be perceived as originating from the *external* location associated with the HRTF.

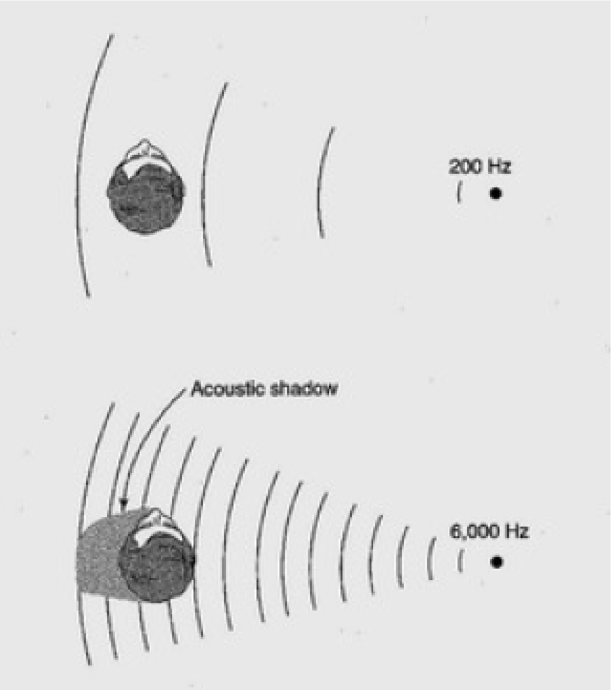


**Figure 6.1.** Differences in IID for a sound source located 90° to an observers' right for low and high frequency stimuli. The low frequency sound diffracts around the observers head, such that minimal differences are apparent in intensity between the two ears. In contrast, the high frequency sound is unable to diffract around the head due to its shorter wavelength and the head therefore causes an 'acoustic shadow'. Due to this shadowing effect, the intensity of sound at the left ear is reduced and thus an IID exists (Goldstein, 1999).

### 6.2.1 Measurement of HRTFs

Initial subjects for the HRTF measurement procedure were author JVMH, JH and NWR. The procedure was performed in a 4m x 6.3m anechoic chamber at the University of Nottingham; the chamber had previously been verified by the Institute of Hearing Research (IHR) at Nottingham as anechoic to sounds above 138Hz. Observers had normal hearing in both ears by self-report (audiometrically verified in the case of observer NWR) and no history of auditory pathology. Recording and analysis of the HRTFs was performed by

NWR, RWD and PVM from the Psychology Department at the University of Nottingham.

A full description of the apparatus, stimuli, recording and off-line analysis used to record the HRTFs is beyond the scope of the present work, and thus the brief description below serves only to illustrate the measurement process. However, a more detailed description of the process used by NWR, RWD and PVM appears in a recently-published article (Deas, Roach, & McGraw, 2008); interested readers are therefore directed to this article for a more detailed account of the HRTF measurement procedure.

Prior to recording, observers were fitted with Electret EK-3133 broadband condenser microphones (Knowles Acoustics, U.S.A.) positioned flush with the entrance to both auditory canals and held in place with Flex AB-40 Shore audiological putty (Egger Otoplastik & Labortechnik, Germany) which completely blocked the auditory canals. Beneath the microphones, foam earplugs ensured that the putty did not damage the internal ear. After being prepared for the measurements in this way, observers sat inside the chamber on a rotatable chair 2m away from a tripod-mounted Acoustimas loudspeaker (Bose, U.S.A.). A laser pointer suspended from the ceiling had previously been aligned with the chair’s centre of rotation, and prior to each measurement the centre of rotation of each individual was also aligned with the laser pointer. Observers were then held in this position with an adjustable chin cup clamped to the chair arm. The tripod was adjusted in height so that the middle of the loudspeaker was horizontally aligned with the auditory canal for each observer.

Auditory stimuli were presented via the speaker, with the responses of the in-ear microphones being recorded by a specially-designed codec box (IHR, Nottingham, U.K.) for further (offline) processing. Measurements were made every 2° between ±90°, with the exception of the central 40° (± 20°) where measurements were made every 1°. In the present work, negative spatial offsets correspond to external events to the right of an observer; thus, for the -90° measurement the speaker was located directly facing the observers' right ear, as in Figure 6.2. In total, recording took approximately two hours for each observer. After this session, recordings were further processed offline to generate the HRTFs corresponding to the spatial offsets ±90° for each observer.

## 6.3 Psychophysical validation of HRTFs

Given the complex nature of HRTF measurement, and the importance of accurate spatial cues in the proposed experiment, each individual performed an auditory lateralisation experiment in order to validate their individual HRTF. If, when convolved with individual HRTFs, stimuli accurately represent spatially discrete, free-field external auditory events, then use of the HRTFs in subsequent experiments would be justified. Any significant lateralisation errors would indicate an error in measurement and/or processing of the

HRTFs and would therefore suggest that they were unsuitable for use in the proposed experiment in that form.

### 6.3.1 Subjects

Author JVMH, JH and NWR participated in the experiment.

### 6.3.2 Methods and stimuli

Auditory stimuli were 200msec bandpass-filtered (200Hz – 12kHz, corresponding to linear emission characteristics of the loudspeakers as measured by IHR, Nottingham) white noise bursts convolved with individual HRTFs corresponding to one of seven physical spatial offsets: -3°, -2°, -1°, 0° (i.e., directly in front of the observer), 1°, 2° and 3°. In the present work, negative spatial offsets correspond to those on the observers' right-hand side. The stimuli were presented over Sennheiser HD650 linear headphones. Stimulus generation and presentation was controlled by custom-written software run in MatLab (Mathworks, U.S.A.) on a Dell desktop PC.

### 6.3.3 Procedures

Observers sat in a darkened room, wearing the headphones and with their eyes closed. Each presentation consisted of a single stimulus (corresponding to one of the seven spatial offsets); observers had to decide whether the stimulus appeared to come from the left or right of midline and respond via a computer keyboard. The precise order of stimulus presentation was determined randomly within a method of constant stimuli. An experimental run contained 20 presentations at each of the seven spatial offsets, and each observer completed three experimental runs making a total of 420 (7 * 20 * 3) presentations per observer. Observers completed two experimental runs prior to data collection for practise purposes, which were excluded from analysis.

For all observers, the proportion of ‘right of midline’ responses was plotted as a function of spatial offset and fitted with a logistic function of the type described in Section 4.3.3.

### 6.3.4 Results

The psychometric function for observer JVMH is shown below as Figure 6.2. At chance performance (50% ‘stimulus right of midline’ responses), the observer is maximally uncertain as to whether the stimulus appeared to be presented right or left of the midline; the physical spatial offset corresponding to this level of performance is therefore the Point of Subjective Equality (PSE; Section 3.1.2). It can be seen that for this observer, the physical and perceived midline co-incide almost exactly.

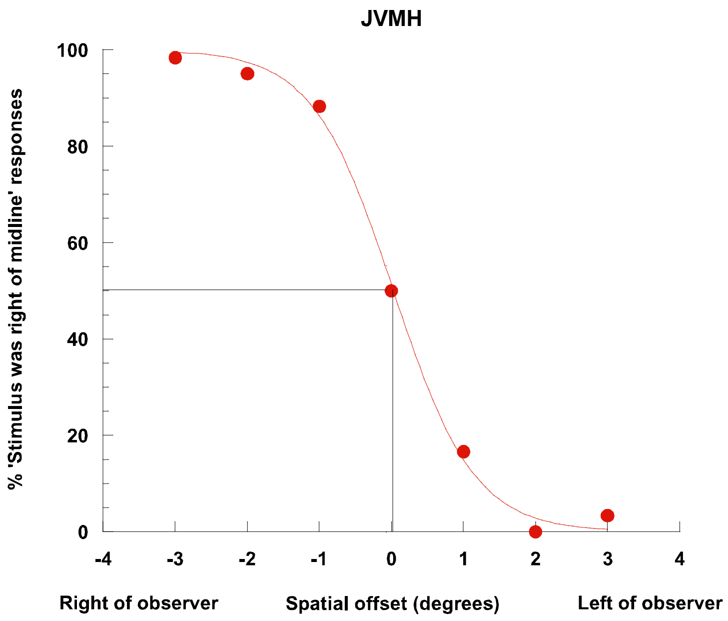


**Figure 6.2.** Psychometric function derived from auditory lateralisation data for observer JVMH. The spatial offset corresponding to 50% 'right-first' responses is the PSE.

Figure 6.3(a & b) shows PSE and JND values for each observer. It can be seen that for all observers, the PSE is approximately zero (Figure 6.3 (a)); this indicates that the HRTFs for each observer contain accurate cues as to the laterality of a sound source. JNDs are approximately 0.5° for each observer.

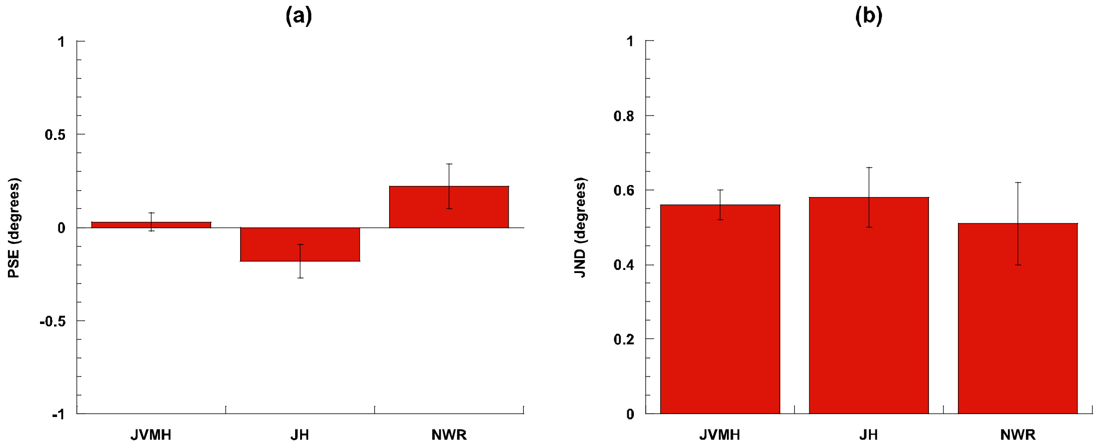


**Figure 6.3(a & b).** PSE values are shown for each observer in Figure 6.3 (a). The sensitivity of observers' auditory lateralisation judgments is shown in Figure 6.3 (b). All values are in degrees. Error bars represent one standard deviation either side of the parameter values.

### 6.3.5 Discussion

For all three observers, performance in the lateralisation task with the HRTF-generated stimuli was approximately veridical, whilst sensitivity to spatial offset is comparable to that measured using free-field stimuli (Perrott & Saberi, 1990). Previous work has indicated that presenting HRTF-derived auditory stimuli over headphones may cause errors in elevation judgments (Wightman & Kistler, 1989b). This was felt unlikely to affect the results in the proposed experiment as all auditory stimuli were constructed from HRTFs measured in the horizontal plane without any elevation cues. The auditory stimuli were thus judged to be suitable for use in the proposed experiment following validation in the lateralisation experiment.

## 6.4 Adaptation to Spatially Disparate Auditory-Visual Lags

### 6.4.1 Subjects

Author JVMH, JH and NWR participated in the experiment. All observers had extensive prior experience of a TOJ task.

### 6.4.2 Methods and Stimuli

The visual stimulus was a Gaussian blob ($\sigma = 2°$, illuminance 100cd/m$^2$) presented for one frame (10msec) -10° (10° right) or 10° (left) of a central fixation cross on a Mitsubishi Diamond Pro 2070 22” CRT monitor (100Hz refresh rate, background illuminance 50cd/m$^2$). The exact moment of presentation was controlled by a ViSaGe Visual Stimulus Generator (VSG) (Cambridge Research Systems, U.K.), which synchronised presentation to the refresh cycle of the monitor to prevent ‘shearing’ of the visual stimulus. The auditory stimulus was a 10msec square-wave windowed burst of bandpass-filtered (200Hz – 12kHz) white noise (70dB SPL) delivered binaurally via Sennheiser HD650 linear headphones. For each auditory presentation, the white noise was convolved with the observers individually recorded HRTF representing the spatial offset being tested (+10° or -10°) to produce an auditory stimulus in spatial register with the corresponding visual stimulus. The experiment was controlled by custom-written software in MatLab (Mathworks, U.S.A.) on a Dell desktop PC.

### 6.4.3 Procedures

Observers participated in an unadapted (baseline) TOJ condition, as well as a total of four adapted TOJ conditions. Adapted conditions were classified as either *congruent* or *incongruent*, according to the relative polarities of the adapting asynchronous stimulus pairs at both locations. In the congruent conditions, observers adapted to either a visual (V) lead 10° right and left of fixation or an auditory (A) lead right and left of fixation. In the incongruent conditions, observers adapted to opposing lags 10° right and left of fixation. The four conditions are illustrated in Figure 6.4(a-d). In all conditions, the visual and auditory components of all stimulus pairs (i.e., of each AV event) were presented at the same spatial location. An experimental run in the baseline condition consisted of 10 AV stimulus pairs presented at each of seven SOAs (-120, -80, -40, 0 40, 80 and 120msec; negative values correspond to a physical lead of sound over vision) at both locations. The order of SOA presentation varied randomly within a method of constant stimuli, and the laterality of the AV events was determined randomly on a trial-by-trial basis.

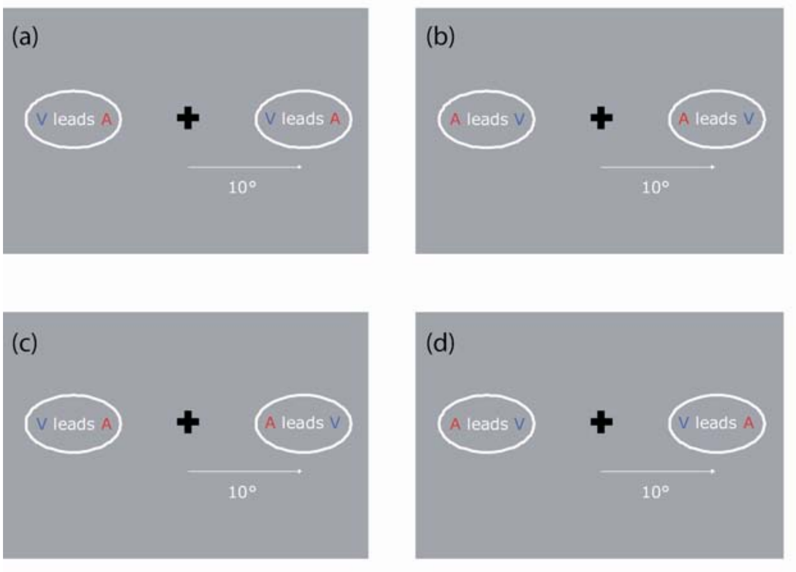


**Figure 6.4(a-d).** Schematic showing the four adaptation conditions employed: adapt visual lead R & L (a), adapt A lead R & L (b), adapt A lead R, A lead L (c), and adapt V lead R, A lead L (d). Hence, conditions (a) and (b) are referred to as *congruent* adaptation conditions, whereas (c) and (d) are referred to as *incongruent* adaptation conditions.

An experimental run in any of the adapted conditions began with an ‘adaptation phase’ consisting of 120 pairs of asynchronous AV stimulus pairs (60 each right and left of fixation); the adapting asynchrony was ±120msec. As in the previous asynchrony adaptation experiments (Sections 5.2.3 and 5.3.2), observers simply attended to the adapting stimulus pairs in the absence of any cognitive task. The laterality of the adapting stimulus pair was determined randomly on a trial-by-trial basis. Each adapting stimulus pair was separated by an interval that varied randomly (with a uniform probability) between 300-600msec. At the end of the adaptation phase, a two-second pause alerted the observer that the ‘test’ phase was about to begin. There was then a top-up phase of six presentations (identical to those in the adaptation phase) alternating either side of fixation, with the seventh presentation being the test pair. The laterality of the test stimulus pair was determined randomly on a trial-by-trial basis. Test pairs were presented at one of seven possible SOAs, as per the baseline condition. Note that in this experiment the range of SOAs tested in each condition was *not* offset, as in Section 5.2.3. The experimental paradigm employed in the adaptation conditions is illustrated in Figure 6.5.

In both the adapted and baseline conditions, observers had to make an unspeeded TOJ as to which of the visual and auditory components of the test stimulus pairing was presented first, and respond via a computer keyboard. Each SOA was presented 10 times within an experimental run, and all observers completed four experimental runs of each condition. This made a total of 560 test presentations (2 stimulus locations * 10 presentations * 7 SOAs * 4 runs) in the baseline condition and each adaptation polarity (vision- or sound-first) for each of the four adaptation conditions, making a final total of 5040 (560 * 9) test presentations per observer. As in previous experiments (Sections 4.3.3 and 5.2.3) observers collected the data for the different conditions in a quasi-random order and switched conditions at the end of each experimental run, in order to eliminate any potential practise effects on analysis of JND values. For each observer, each condition and each adaptation polarity the percentage of 'sound-first' responses was plotted as a function of SOA and fitted with a logistic function identical to that described previously (Section 4.3.3).

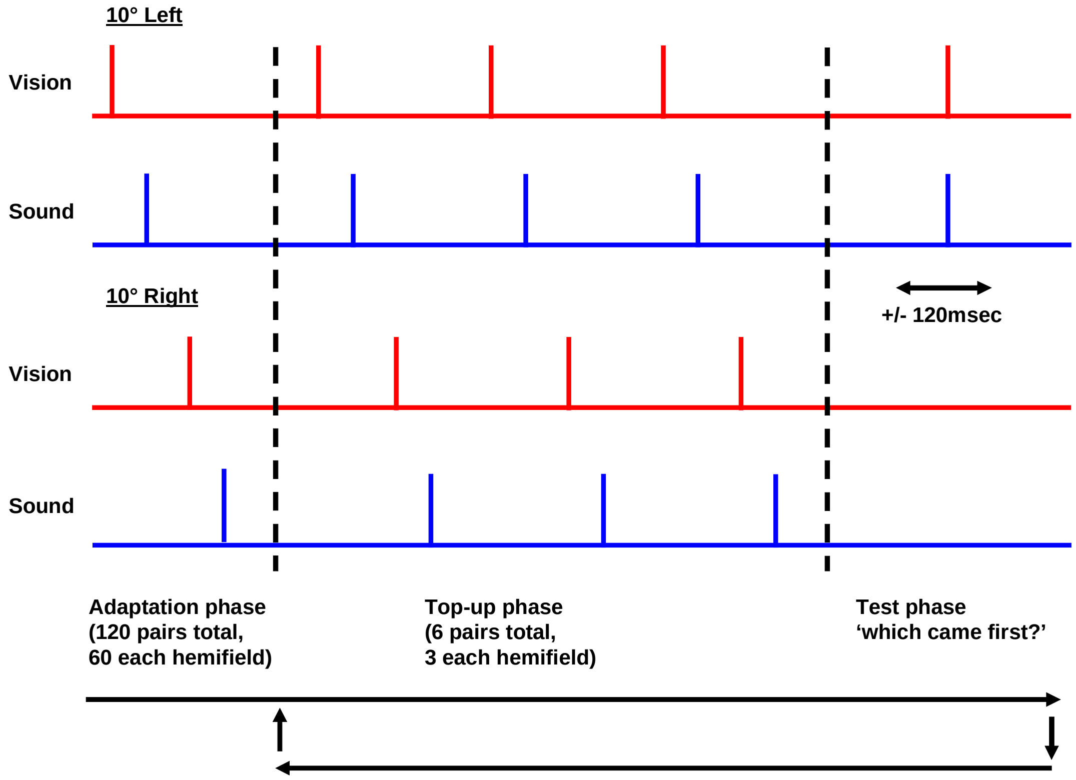
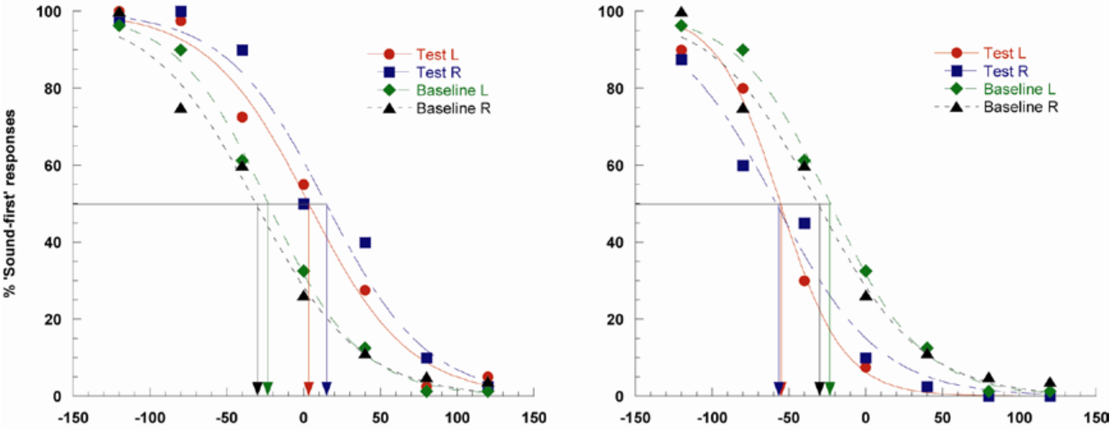
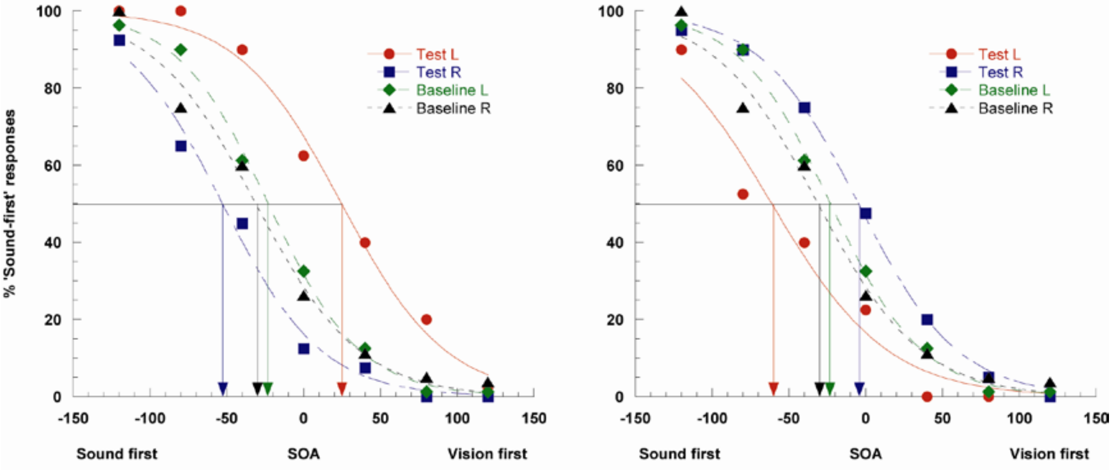


**Figure 6.5.** Illustration of the paradigm used in the adaptation conditions. Observers were exposed to 120 pairs of asynchronous AV stimulus pairs (60 each 10° right and left of fixation; in the example shown, 'vision leads sound' right and left of fixation). Following the adaptation phase, a further six adapting stimulus pairs were presented (three each either side of fixation) followed by a test pair presented at only one location (determined randomly on a trial-by-trial basis; in the example shown, left of fixation). Observers judged whether the auditory or visual stimulus appeared to be presented first and responded via the computer keyboard. The response triggered the next cycle of top-up and test stimuli until the experimental run was completed.

### 6.4.4 Results


(a) congruent - adapt V lead R & L

(b) congruent - adapt A lead R & L


(c) incongruent - adapt V lead L, A lead R

(d) incongruent - adapt A lead L, V lead R

**Figure 6.6(a-d).** Psychometric functions for observer JVMH following adaptation to asynchronous AV stimulus pairs 10° right and left of fixation. The polarities of the adapting stimulus pairs were either congruent (a-b) or incongruent (c-d) across locations; the four graphs correspond to the conditions illustrated in Figure 6.4(a-d). In every case, the physical SOA corresponding to 50% 'sound-first' responses is the Point of Subjective Simultaneity (PSS). Baseline functions are identical on all four graphs.

Psychometric functions for observer JVMH are shown as Figure 6.6(a-d, corresponding to the schematic in Figure 6.4). On each graph, the data points pertaining to the baseline (unadapted) TOJs are represented by green diamonds and black triangles (relating to TOJs made 10° left and right of fixation, respectively); these data points and corresponding psychometric functions are therefore identical on all four graphs.  Examination of the functions shows that following adaptation to asynchronous AV stimulus pairs presented 10° right and left of fixation, PSS is shifted in the direction of the adapting lag.  This is the case in both the congruent (Figure 6.6(a–b)) and incongruent (Figure 6.6(c–d)) adaptation conditions.  For example, adapting to a visual lead 10° left of fixation and an auditory lead 10° right of fixation causes PSS to correspond to a physical lead of vision over sound left of fixation, and an increased lead of sound over vision right of fixation (Figure 6.6(c)).

For each observer, the magnitude of the adaptive shifts in PSS (defined as the difference between adapted and unadapted PSS values at each location in each condition) was plotted, as shown in Figure 6.7(a-d).  As in Figure 6.6, conditions a-d correspond to those illustrated in Figure 6.4(a-d).  The figure shows that the pattern of temporal recalibration effects shown by author JVMH closely matches that of the other observers in both the congruent (a-b) and incongruent (c-d) adaptation conditions.

ANOVA revealed no significant differences in the magnitude of PSS shifts between the congruent and incongruent conditions ($F_{1, 2} = 2.04$, $p > 0.05$) or between the left and right hemifields ($F_{1, 2} = 0.06$, $p > 0.05$).  Although it could be argued that this may reflect the reduced statistical power of ANOVA on a relatively small pool of observers, it is consistent with the effects shown in Figures 6.6(a-d) and 6.7(a-d). As expected, the effect of adaptation polarity was highly significant ($F_{1, 2} = 117.0$, $p < 0.01$), confirming that adaptation to ‘vision leads sound’ causes PSS to correspond to an increased physical lead of vision over sound, and vice versa.

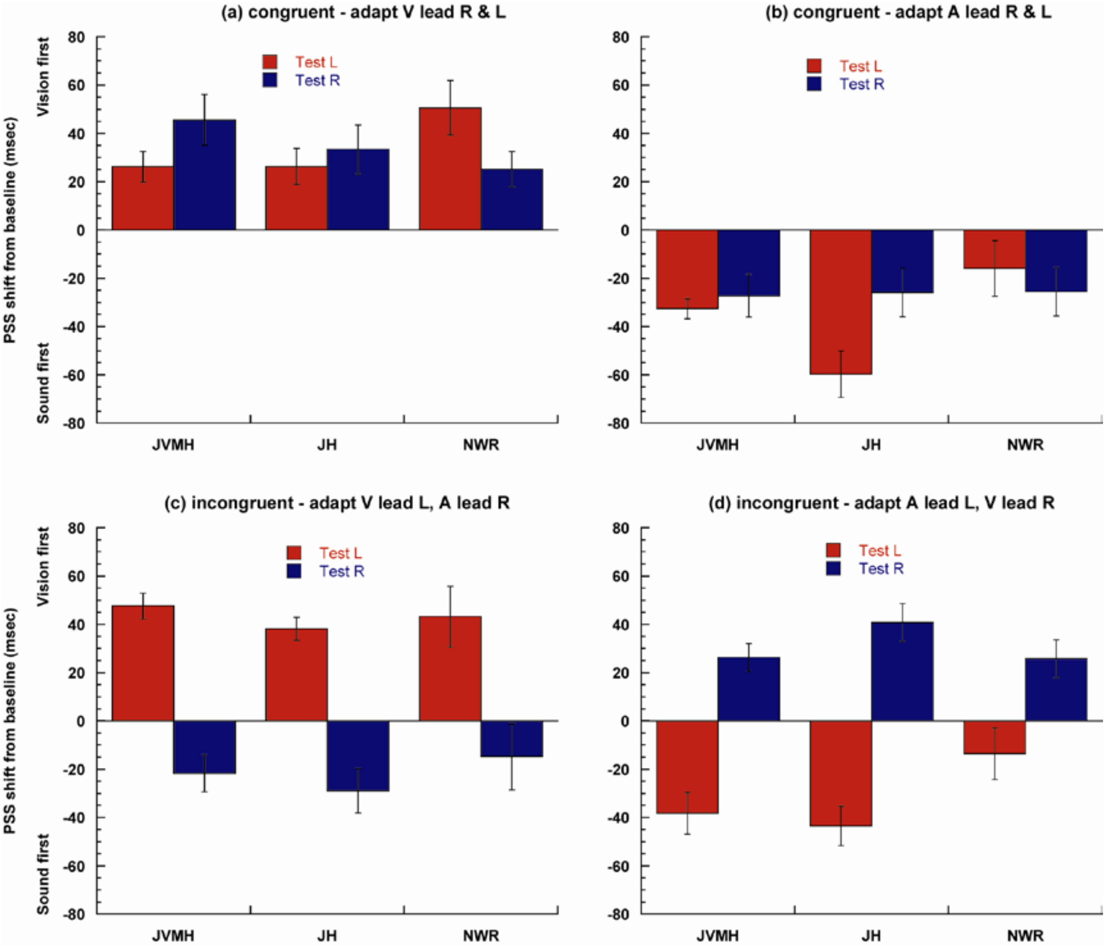


**Figure 6.7(a-d).** PSS shifts from baseline for all observers in the congruent (a-b) and incongruent (c-d) adaptation conditions 10° right and left of fixation (red and blue bars, respectively). The four graphs correspond to the conditions depicted in Figure 6.5(a-d). Error bars represent one standard deviation either side of the parameter values.

An additional repeated-measures ANOVA comparing JNDs in all conditions (including baseline) and locations revealed no significant differences in observer sensitivity ($F_{2, 18} = 1.06$, $p > 0.05$) between the different conditions. In other words, JND was unaffected by the asynchrony adaptation process, test location, and adapting asynchrony polarity.

### 6.4.5 Discussion

The results demonstrate for the first time that perceived AV temporal order may be modulated by adaptation to asynchronous stimulus pairs in a spatially-specific manner. This temporal recalibration appears to be similar in nature in the congruent and incongruent adaptation conditions, as evidenced by the close correspondence in both PSS shifts and JND values. The temporal recalibration effects in the congruent conditions are similar in nature to those presented earlier this thesis (Sections 5.2.4 and 5.3.3) and elsewhere (Fujisaki et al., 2004; Vroomen et al., 2004). However, the findings in the incongruent adaptation conditions (adaptation to opposing AV lags at different locations produces correspondingly opposing PSS shifts at both locations) are perhaps surprising.

Following completion of the present experiment, Keetels and Vroomen (2007) published data demonstrating that asynchrony adaptation effects are not specific to the *relative* spatial locations of the auditory and visual stimuli. In this case, the apparatus used was a LED and two loudspeakers, one of which was spatially co-localised with the LED and the other offset laterally by 90°. The four experimental conditions used are illustrated in Figure 6.8; results showed no significant differences between the temporal recalibration effects in the four conditions (Keetels & Vroomen, 2007). This result is more consistent with the hypothesis that asynchrony adaptation effects are non-spatially specific, and may appear to argue against the results of the present experiment.

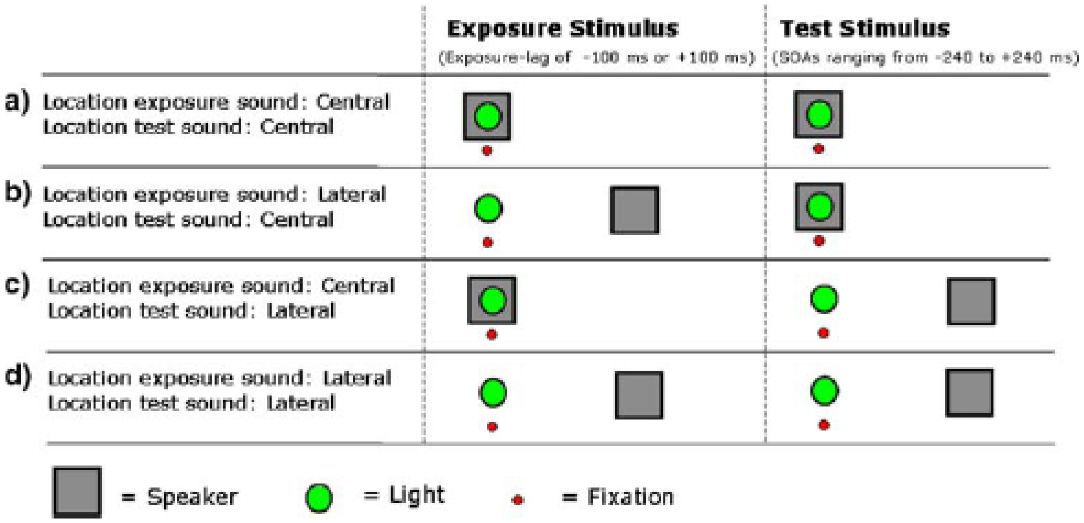


**Figure 6.8.** Experimental conditions employed in a previous examination of spatial factors in asynchrony adaptation (Keetels & Vroomen, 2007). Apparatus used was a green LED located adjacent to fixation, and two loudspeakers (one located in spatial alignment with the LED, the other 90° lateral). In the adaptation phase, where observers were exposed to asynchronous AV stimulus pairs (‘Exposure Stimulus’), visual and auditory stimuli were presented either in spatial register (conditions (a) and c)) or 90° apart (conditions (b) and d)). Stimuli in the test phase were also presented either spatially co-localised (conditions (a) and (b)) or 90° apart (conditions (c) and (d)). The magnitude of PSS shifts was found not to differ significantly in the four conditions (Keetels & Vroomen, 2007).

The data presented here may also be considered surprising in light of the asynchrony adaptation effects recorded elsewhere in this thesis (Sections 5.2.4 and 5.3.3) showing similar temporal recalibration effects and JNDs in the AV, AT and VT modality pairings, consistent with a late-stage recalibration mechanism. Fujisaki et al. (2004) also presented results suggestive of a relatively late-stage temporal recalibration mechanism (i.e., temporal recalibration was not dependent on low-level stimulus features such as auditory stimulus laterality being identical in the adaptation and test phases).

However, it is also possible that both these previous findings are complementary to those of the present study. Considering first the findings of Keetels and Vroomen (2007) suggesting a unitary recalibration mechanism which generalises over space, a critical difference with the present work is that only a *single* adaptive state was induced in observers at any one time (compared to two in the present work). It is therefore possible that when exposed to a single instance of asynchronous AV stimulus pairs, the temporal aftereffects are evident over the entire region of external space (or at least over approximately 90° (Keetels & Vroomen, 2007)). Conversely, concurrent adaptation to two instances of spatially-distinct asynchronous AV stimulus pairs may allow the spatially-specific effects described here to be invoked. The fact that the PSS shifts recorded here (27% of the adapting lag, on average) are comparable with those recorded following adaptation to a single pair of AV (and also AT and VT) stimulus pairs (Sections 5.2.4 and 5.3.3) may suggest that the effects are in fact mediated by the same underlying mechanism in both instances.

Whilst the spatial specificity of the effects demonstrated here is certainly compatible with an early-stage cortical or subcortical recalibration mechanism, another alternative exists. The spatial specificity may be attributable to a high-level, cognitive ‘association’ built up between the stimulus pairs at both locations. For example, during the adaptation phase the cognitive system may build up associations between the two stimulus pairings and the temporal relationship between the individual stimuli at both locations. Following the acquisition of such experience, the asynchrony adaptation mechanism may then allow temporal recalibration at *both* locations. Importantly, such an explanation would imply that the results of the present experiment are not spatially-specific *per se*, but rather that the perceptual system is able to maintain multiple adaptive states concurrently as long as the multiple adapt and test stimulus pairs can be clearly differentiated (in this case, by spatial location). This hypothesis is much easier to reconcile with data presented within this thesis (Sections 5.2.4 and 5.3.3) and the findings of Fujisaki et al. (2004) discussed above in that it avoids recourse to an early-stage recalibration mechanism, which would be sensitive to low-level stimulus features such as auditory stimulus laterality.

Such a hypothesis could be tested in at least two ways. Firstly, if the results of the present experiment reflect a genuinely late-stage, cognitive, mechanism, then the number of adaptive states that an observer can concurrently maintain whilst still displaying adaptive temporal recalibration should be limited by the available cognitive resources. Conversely, if the effects reflect multiple low-level spatially-specific recalibration mechanisms, then the potential number of concurrent adaptive states would be limited only by the spatial specificity of such mechanisms – in theory, a much larger number. However, given the large number of adapt, test and top-up trials necessary in the current experiment, repeating the experiment with observers adapting to different lags at more than two locations presents considerable methodological challenges.

An alternate way to test this hypothesis would be to differentiate the different stimulus pairings by some factor other than spatial location. For example, visual stimuli could be defined by chromaticity (e.g., equiluminant red vs. green) and auditory stimuli by pitch (e.g., high vs. low). Observers could then adapt to two sets of asynchronous AV stimulus pairs with arbitrary colour/pitch combinations at the same location, and make TOJs using the same stimulus pairs at the same location. An example of one of the incongruent adaptation conditions possible in such an experiment is shown as Figure 6.9. If asynchrony adaptation effects are genuinely spatially-specific in nature, then it is unlikely that any temporal recalibration would be observed in the incongruent conditions due to the two stimulus pairs being spatially co-localised and the incongruent adaptation polarities (vision leads sound vs. sound leads vision) negating each other. However, a high-level associative mechanism would permit adaptive PSS shifts in both stimulus pairings. In the example shown in Figure 6.9, the PSS of a green visual and low-pitched auditory stimulus pair would become more ‘vision-first’, whilst that of a red visual and high-pitched auditory stimulus pair would become more ‘sound-first’. Stimulus pairs could be differentiated by any number of potential features; for example, visual stimuli could also be differentiated by their orientation, spatial frequency or shape.

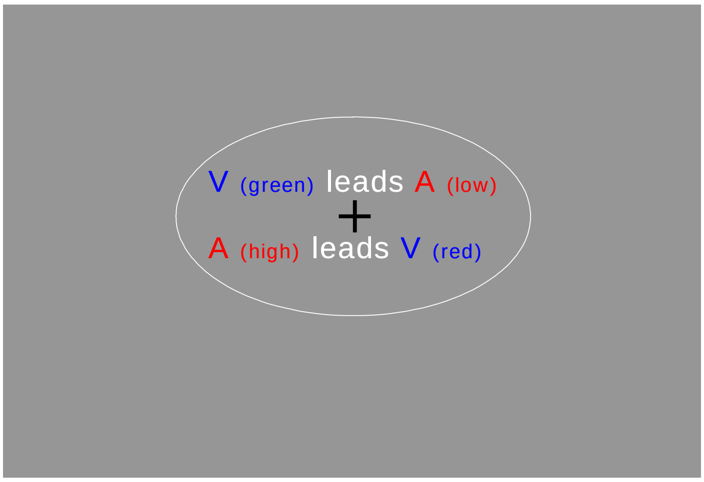


**Figure 6.9.** Example of potential adaptation conditions when stimulus pairs are no longer defined by location. In this example using two stimulus pairings, all stimuli are presented at fixation but a green visual stimulus is paired with a low-pitched auditory stimulus (vision leads sound), whilst an equiluminant red visual stimulus is paired with a high-pitched auditory stimulus (sound leads vision). Following a period where observers adapted to both the ‘vision leads sound’ and ‘sound leads vision’ stimulus pairs, TOJs would be made at fixation with the same stimuli.

Other features of the results are also worthy of comment. Firstly, the observed PSS shifts are recorded in the absence of any loss of observer sensitivity to asynchrony. This confirms the results described in Section 5.2.4 and recent literature (Harrar & Harris, 2008; Heron et al., 2007). In addition, the significant shifts in AV PSS measured using a method of constant stimuli *without* an offset range of SOAs provide further evidence against the potential criticism that the temporal recalibration previously reported (Section 5.2.4) may be ascribable to an offset SOA range and response bias combined with the method of constant stimuli. Finally, the observed PSS shifts were elicited with only 60 adapting stimulus pairs at either location (rather than 100 in Sections 5.2.4 and 5.3.3), and only three ‘top-up’ stimulus pairs at either location (rather than five in Sections 5.2.4 and 5.3.3). Although the precise temporal dynamics of the asynchrony adaptation process are still unknown, it is clear from these results that asynchrony adaptation may be a more rapid process than is apparent from previous work (Fujisaki et al., 2004; Harrar & Harris, 2008; Keetels & Vroomen, 2007; Vroomen et al., 2004).

## 6.5 Influence of attention on asynchrony adaptation and temporal order judgments

### 6.5.1 Introduction

Although the AV temporal recalibration effects recorded within this thesis (Sections 5.2.4, 5.3.3 and 6.4.4) are qualitatively similar to those recorded in previous work (Fujisaki et al., 2004; Harrar & Harris, 2008; Keetels & Vroomen, 2007; Vroomen et al., 2004), quantitative differences in the magnitude of temporal recalibration are apparent between studies. For example, within this thesis an average AV temporal shift (defined here as the difference between PSS values measured following adaptation to 'vision leads sound' and 'sound leads vision' asynchronous stimulus pairs) of 73msec (40.5% of the adapting lag) was recorded (Section 5.2.4), whilst Keetels and Vroomen (2007) recorded average temporal shifts of only 12.9msec (6.5% of the adapting lag). Currently, the factor(s) that may modulate temporal recalibration subsequent to asynchrony adaptation are unstudied. As well as potentially providing an explanation for the differing effect sizes between studies, any such factors are of interest because they may yield clues as to the nature of the neural mechanism mediating temporal recalibration. Currently, the weight of evidence suggests a relatively late-stage neural substrate (Sections 5.2.4 and 5.3.3; Fujisaki et al., 2004); however results presented in Section 6.4.4 may be consistent with an early-stage mechanism specific to the adapted region of external space.

A notable difference between the experiments presented in this thesis and elsewhere (Fujisaki et al., 2004; Harrar & Harris, 2008; Keetels & Vroomen, 2007; Vroomen et al., 2004) is that in the present work, no secondary task was performed by observers during the adaptation phase; rather, observers simply attended to the adapting stimulus pairs without being required to make any judgment of the stimuli. In contrast, Vroomen et al. (2004) required observers to detect 'oddball' visual stimuli (of different chromaticity to the adapting visual stimuli) presented at fixation, and Keetels and Vroomen (2007) required participants to detect the offset of a fixation light. Similarly, Fujisaki et al. (2004) required observers to detect larger visual stimuli and auditory stimuli of different pitch. Thus, observers' attention was directed to low-level, *non-temporal* features of the adapting stimuli. It is currently unknown whether observers' attention during the adaptation phase is an important variable in adaptive temporal recalibration. However, significant effects of attention on a range of timing tasks have been recorded in the absence of adaptation (Chaston & Kingstone, 2004; Correa, Sanabria, Spence, Tudela, & Lupianez, 2006; Rolke, Ulrich, & Bausenhart, 2006; Spence et al., 2001b; Tse et al., 2004; Zampini et al., 2005c).

In the non-temporal domain, the available evidence strongly suggests that observers' attention may modulate sensory aftereffects. The motion aftereffect (MAE) is an illusion where observers exposed to one direction of visual motion (e.g., by viewing a waterfall) perceive a subsequent stationary scene as moving in the opposite direction to that experienced during adaptation (Mather, Verstraten, & Anstis, 1998; Sekuler & Blake, 1994). Previous work has found that when observers monitored a stream of letters at or adjacent to fixation to detect oddball presentations of numerals, the duration of the MAE was significantly reduced relative to a control condition where observers simply used the stream of alphanumeric characters as a fixation point (Chaudhuri, 1990). Critically, the two conditions were identical in every respect other than that observers' attentional resources were employed by the secondary task only in the experimental condition. Thus, a cognitive task unrelated to the adapting stimuli or subsequent judgment of the test stimuli appears to modulate the magnitude of the illusion. A similar result was obtained when the secondary task employed was auditory (rather than visual) in nature (Zhou & Chen, 1994).

Subsequently, it has been demonstrated that attending to a single *direction* of motion in a display containing superimposed and opposing motion stimuli increased the chances of observers perceiving a subsequent dynamic noise pattern as moving in the opposite direction to that attended to, i.e. increased the likelihood of observers experiencing a MAE specific to the focus of attention during adaptation (Lankheet & Verstraten, 1995). Further, it has been reported that the MAE was enhanced when observers made motion speed judgments of the adaptation pattern relative to when they made judgments about its luminance or colour (Boutet, Rivest, & Intriligator, 1996).

Together, these findings suggest that the use of a distracter task may impair observers' perception of the MAE, but that when the task requires monitoring the parameter judged during the test phase, the MAE may be enhanced. Attention also appears to modulate the tilt (Spivey & Spirn, 2000) and stereoscopic depth (Rose, Bradshaw, & Hibbard, 2003) aftereffects, demonstrating that the effect of attention is not unique to the MAE.

Given that attention appears to modulate performance in timing tasks and the perception of sensory aftereffects, the possibility arises that it may also modulate the magnitude of asynchrony adaptation effects. Previous studies offer support for this hypothesis. For example, the study of Keetels and Vroomen (2007) recorded temporal shifts of 12.9msec (6.7% of the adapting lags). Observers in this study were required to detect and respond to the offset of a fixation light during the adaptation phase. In contrast, Harrar and Harris (2008) asked their observers to "…pay attention to the (adapting stimulus) pairs either by counting them or by trying to decide their temporal order". This study found that PSS measured after observers adapted to a physical lead of vision over sound shifted by 32msec (32% of the adapting lag) relative to an unadapted baseline. This suggests that temporal recalibration may be greatest when observers are given the option to attend to the temporal order of the adapting stimuli.

With these factors in mind, a task was devised to investigate the effect of attention on AV TOJs made subsequent to asynchrony adaptation. It was hoped that the results would offer further clues as to the nature of the neural substrate mediating adaptive temporal recalibration. If AV asynchrony adaptation effects are modulated by the focus of observers' attention, as discussed above, this would suggest that the mechanism(s) mediating asynchrony adaptation is/are relatively late-stage and post-attentive in nature.

Conversely, should attention not modulate the magnitude of asynchrony adaptation effects, the notion of a low-level, pre-attentive mechanism is more credible. This outcome would also support the possibility that the concurrent opposing recalibration effects recorded in Section 6.4.4 are most likely mediated at an early, pre-attentive stage of the perceptual system (and thus genuinely spatially-specific in nature), rather than reflecting a high-level associative mechanism as suggested by the results in Sections 5.2.4 and 5.3.3.

### 6.5.2 Subjects

Observers were author JVMH, JH, NWR, and naïve observers AA and MB. All had extensive prior experience of a TOJ task. Prior to the experiment, observers AA and MB had their individual HRTFs measured and psychophysically validated in an identical manner to that described previously (Section 6.3.3).

### 6.5.3 Methods and stimuli

The visual stimulus was a Gaussian blob (σ = 2°, illuminance 100cd/m2) presented for two frames (20msec) at the centre of a Mitsubishi Diamond Pro 2070 22” CRT monitor (100Hz refresh rate, background illuminance 50cd/m$^2$). The exact moment of presentation was controlled by a ViSaGe Visual Stimulus Generator (VSG) (Cambridge Research Systems, U.K.), which synchronised presentation to the refresh cycle of the monitor to prevent ‘shearing’ of the visual stimulus. The centre of the stimulus was aligned with the centre of a fixation cross. The auditory stimulus was a 20msec square-wave windowed burst of bandpass-filtered (200Hz – 12kHz) white noise (70dB SPL) delivered binaurally via Sennheiser HD650 linear headphones. For each auditory presentation, the white noise was convolved with the observers’ individually-recorded HRTF representing 0° azimuth to produce an auditory stimulus spatially coincident with the visual stimulus. The experiment was controlled by custom-written software in MatLab (Mathworks, U.S.A.) on a Dell desktop PC.

### 6.5.4 Procedures

At the start of each experimental run, observers fixated the central cross on the computer monitor and pressed a key when ready to begin the experiment. Each experimental run began with an ‘adaptation phase’, during which observers were exposed to 120 AV stimulus pairs with a uniform SOA of 120msec. The polarity of the adapting stimulus pairs (e.g., ‘vision leads sound’) was consistent throughout an experimental run. Each stimulus pair was separated by an ISI which varied randomly (with a uniform probability) between 200-400msec. Following the adaptation phase, a pause of two seconds alerted the observer that the ‘test phase’ of the experiment was about to commence. Before each test presentation, observers were presented with four ‘top-up’ stimulus pairs of the same polarity as in the adaptation phase, with the fifth stimulus pair being the test stimulus. The test stimuli were presented at one of seven possible SOAs: -120, -80, -40, 0 (simultaneous), 40, 80 and 120msec, which were randomly interleaved within a method of constant stimuli. In the present study, positive SOAs refer to a physical lead of vision over sound. Observers made an unspeeded TOJ of the test stimulus pair as to ‘which modality came first’ and responded via the computer keyboard.

Throughout the course of an experimental run, ‘oddball’ stimuli perceptually distinct from the regular adapt and top-up stimulus pairs were presented. The probability of each oddball occurring on any given adaptation or top-up presentation was 5%. Observers had to respond as quickly as possible upon detecting an oddball stimulus by pressing a key on a computer keyboard with the same hand used to respond to TOJ trials. The time taken for observers to respond to each oddball presentation (i.e., the reaction time (RT)) was recorded for off-line analysis. When an oddball stimulus pair was presented during the adaptation phase, that trial was excluded from the total of 120 adapting stimulus pairs. However, when an oddball was presented during the top-up phase the top-up cycle was reset, such that following observers' detection of the oddball stimulus the cycle of four top-up stimulus pairs followed by the test pairing was re-initiated. Three different oddball stimulus pairings were used:

- 'Attend fixation' condition: the contrast polarity of the central fixation cross was reversed. This was apparent to observers as a change in the colour of the cross from black to white.

- 'Attend stimuli' condition: the visual stimulus was presented at half of its regular size ($\sigma = 1°$) and the auditory stimulus was convolved with the HRTF corresponding to a spatial offset of +10° (i.e., 10 degrees to the left of each observers' midline).

- 'Attend temporal order' condition: the temporal polarity of the adapting stimulus pairs was reversed, such that (e.g.) during adaptation to a 120msec physical lead of sound over vision, the oddball stimulus pair consisted of a visual stimulus physically leading an auditory stimulus by 120msec. Thus, observers had to detect a change in SOA of 240msec.

Prior to commencing each experimental run, observers were instructed that the three types of oddball stimuli were equally likely to be presented during the course of an experimental run, but that they should respond only to the target oddball for that run. The critical feature of the paradigm was that experimental runs in each of the three conditions were identical in all respects other than the focus of observers' attention, which was directed to a different aspect of the AV stimulus pairs in each condition.

Each experimental run contained 10 presentations at each SOA and each observer completed three experimental runs in each of the three oddball conditions, making a total of 1260 presentations per observer (10 presentations * 7 SOAs * 3 experimental runs * 2 adaptation polarities * 3 oddball conditions). Observers performed the different oddball conditions sequentially, with the condition changing after each experimental run, in order to minimise any possible practise effects on JND.

For all observers, the percentage of 'sound-first' responses for each condition was plotted as a function of SOA and fitted with a logistic function identical to that described previously (Section 4.3.3).

### 6.5.5 Results

Psychometric functions for observer JH are shown in Figure 6.10(a). Inspection of the functions suggests that the magnitude of temporal recalibration (as indicated by the lateral distance between the midpoints of the two functions for each experimental condition) is largest in the 'attend temporal order' ('attend TO') condition (blue data points/functions). Median RT for oddball stimulus pairs appears to be comparable in the three conditions for this observer (Figure 6.10(b)). This suggests that the three oddball detection tasks are of comparable difficulty. The decision to use the median (rather than mean) RT as measure of central tendency was taken following inspection of the raw RT data for all observers; this revealed that a minority of oddball RTs were much longer than the majority of values. This was felt likely to reflect occasional lapses in observers' attention during the extended (~20minutes) experimental runs. Median RT was therefore used to minimise the influence of these outliers on RT values.

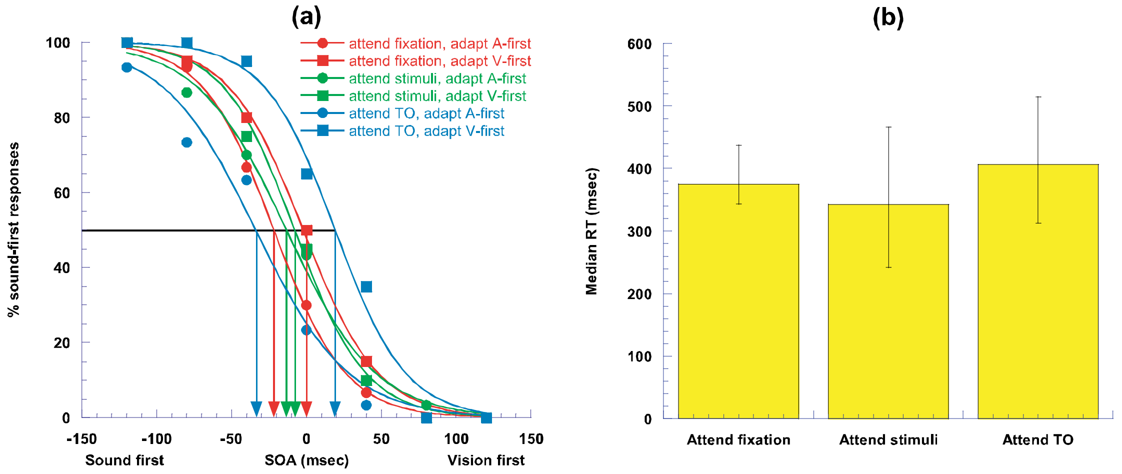


**Figure 6.10(a & b).** Figure 6.10(a) shows psychometric functions for observer JH following adaptation to 'sound leads vision' (circles) and 'vision leads sound' (squares) asynchronous AV stimulus pairs for the attend fixation (red points/curves), attend stimuli (green points/curves) and attend temporal order (TO; blue curves/functions). The difference between the mid-points of each pair of functions indicates the magnitude of the temporal recalibration effect for that condition. Figure 6.10(b) shows median RT to the oddball stimulus pairs in each of the three conditions for observer JH. Error bars indicate the interquartile range for each condition.

The magnitude of the temporal recalibration effect for each observer in each condition was calculated as the algebraic difference between PSS values in the ‘adapt vision leads sound’ and ‘adapt sound leads vision’ conditions[10]. This is shown averaged across observers in Figure 6.11(a). The pattern of results displayed by observer JH is typical of the group as a whole: temporal recalibration appears to be greatest when observers attend to the temporal relationship between the individual visual and auditory stimuli which comprise the adapting and top-up AV stimulus pairs. ANOVA confirmed that the temporal recalibration effects differed significantly in magnitude between the three conditions ($F_{2, 8} = 18.71$, $p < 0.001$). Post-hoc analysis corrected for multiple comparisons (Tukeys HSD) revealed that the ‘attend temporal order’ effects differed significantly in magnitude from both the ‘attend fixation’ and ‘attend stimuli’ conditions ($p < 0.005$). In contrast, no significant differences were found between the ‘attend fixation’ and ‘attend stimuli’ conditions ($p > 0.05$).

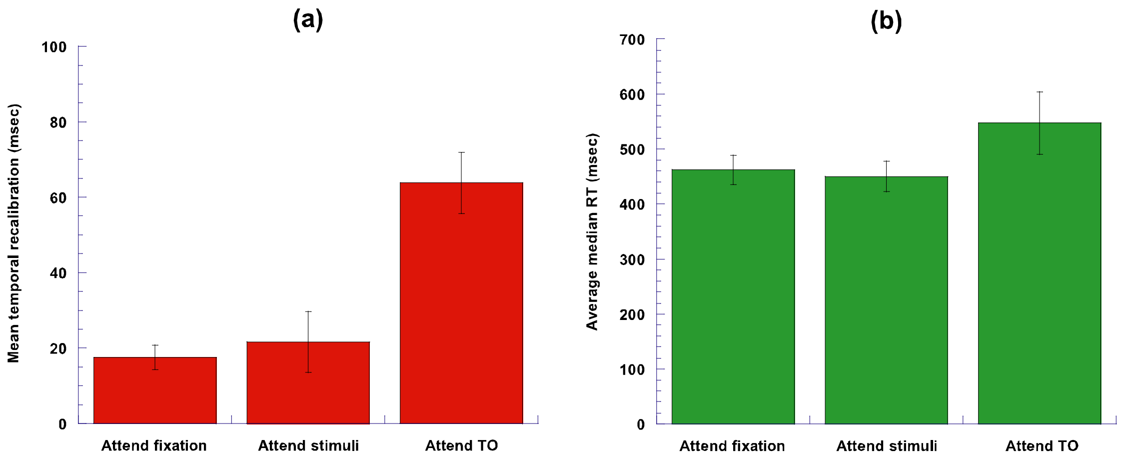


**Figure 6.11(a & b).** Figure 6.11(a) shows the mean temporal recalibration effect (n = 5) as a function of oddball detection task. Figure 6.11(b) shows average median RT in the three different conditions. Error bars represent one SEM either side of the parameter values.

---

[10] Additional analysis of PSS shifts from baseline was considered unnecessary, given that previous results showed robust shifts in AV PSS from an unadapted baseline following asynchrony adaptation (Sections 5.2.4, 5.3.3 and 6.4.4).

Analysis of median oddball detection RTs (Figure 6.11(b)) revealed no significant effect of condition ($F_{2,\ 8} = 3.75$, $p > 0.05$). This shows that the pattern of RT values exhibited by observer JH (Figure 6.10(b)) was consistent across observers. Thus, the differences in temporal recalibration between the three conditions cannot simply be ascribed to inter-condition differences in oddball detection task difficulty or attentional demand as evidenced by oddball detection RT. Finally, mean JND values for all observers are summarised according to condition in Figure 6.12. Analysis of JND values found no significant effects of adaptation polarity ($F_{1,\ 4} = 7.21$, $p > 0.05$), oddball detection task ($F_{2,\ 8} = 0.82$, $p > 0.05$), and no significant interaction between these factors ($F_{2,\ 8} = 0.05$, $p > 0.05$).

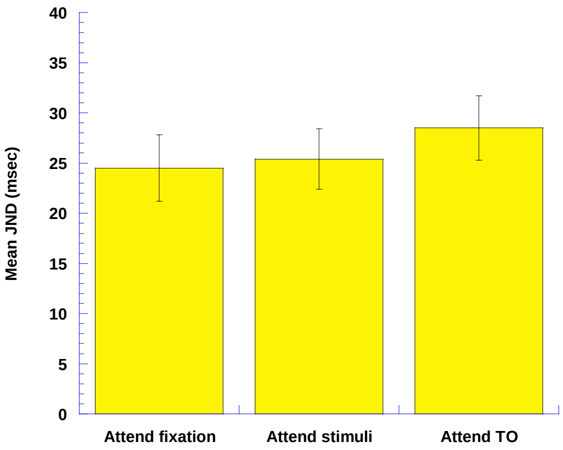


**Figure 6.12.** Mean JND in the three conditions pooled across adaptation polarities. Error bars represent one SEM either side of the parameter values.

### 6.5.6 Discussion

The results of the present experiment demonstrate, for the first time, that the process of AV asynchrony adaptation can be significantly modulated by attention. When observers attend to the temporal order of the adapting stimulus pairs, the total recalibration effects are significantly larger (on average 64msec, or 27% of the adapting lag) than when attention is centred on non-temporal stimulus features (‘attend stimuli’ condition) or is drawn away from the adapting stimulus pairs completely (‘attend fixation’ condition). Both these latter conditions produced an average temporal recalibration of 18-22msec (approximately 8% of the adapting lag) (Figure 6.11(a)). Crucially, these effects were recorded without any statistically significant differences in the difficulty of the oddball task (as inferred from RT) between the three conditions (Figure 6.11(b)). Thus, it is difficult to attribute the pattern of results to inter-task differences in the amount of attention required to perform the oddball task; rather, the *focus* of attention, rather than the degree of attention required to perform the oddball task, appears to be the critical parameter.

Additionally, the lack of any effect of task on JND (Figure 6.12) shows that the difference in recalibration cannot be ascribed to increased temporal uncertainty in the ‘attend temporal order’ condition. Given the inter-condition similarities in JND, it is likely that observers’ attention in the adapt and top-up phases does not affect observer sensitivity to AV asynchrony.

The finding that temporal recalibration is modulated by attention provides further support for the hypothesis that TOJs subsequent to asynchrony adaptation are mediated by a late-stage (post-attentive) neural substrate. This result alone does not confirm that asynchrony adaptation effects are mediated at a late stage of the perceptual system, as attentional modulation of some aftereffects is thought to occur via top-down influences on early, pre-attentive neural processes (Rose et al., 2003; Spivey & Spirn, 2000). However, interpreting attentional modulation of asynchrony adaptation as reflecting a late-stage substrate is justified given that such a mechanism is consistent with other results reported both within this thesis (Sections 5.2.4 and 5.3.3) and elsewhere (summarised in Sections 5.1.1 and 5.3.4). The results also render less likely the possibility that the concurrent opposing temporal recalibration recorded in Section 6.4.4 reflects an early-stage, pre-attentive mechanism.

It can also be inferred from the results that when observers are not given specific instructions prior to exposure to a series of asynchronous stimulus pairs, they may unconsciously direct their attention to the temporal relationship and/or temporal order of the adapting stimulus pairs. This is suggested by the correspondence between the magnitude of recalibration in the ‘attend temporal order’ condition and previous work (Sections 5.2.4 and 5.3.3; Harrar and Harris, 2008). It is noteworthy that the present study is the first to explicitly instruct observers to attend to the temporal relationship between the adapting stimuli; Harrar and Harris (2008) instead gave observers the option to attend to the temporal order or simply count the adapting stimulus pairs. The strong similarity in recalibration effects between this previous study and the present work may imply that previous observers (Harrar & Harris, 2008) chose (consciously or subconsciously) to attend to the temporal order of the adapting stimuli. This does not necessarily imply that the temporal relationship between stimuli promotes exogenous attentional capture under all circumstances; a further possibility is that temporal aspects of the stimuli may attract observers' attention only when observers are aware that the task requires a temporal judgment (as in the work of Harrar and Harris (2008)) or when observers are instructed to focus attention on the temporal order of the stimuli, as in the present work.

Future work will enable a greater understanding of the effects observed in the present experiment. In particular, the conclusion that the focus of observers' attention (rather than secondary task difficulty) is critical to the magnitude of temporal recalibration could be investigated further by repeating the 'attend temporal order' condition with a range of oddball SOAs. It is conceivable that different degrees of attentional focus were exerted to maintain similar RT values for the three different oddball conditions. If RTs of approximately 450-500msec required more attentional resources when the oddball is of a temporal rather than spatial or contrast-based nature, differences in PSS shifts may simply reflect differing allocation of attention across the conditions. Changing the salience of the temporal order oddball would allow quantification of the relationship between RT performance, attentional resource allocation and shifts in PSS. Temporal recalibration effects remaining constant across different oddball SOAs whilst RT varies would provide additional evidence that the object of observers' attention, rather than oddball task difficulty, is the critical parameter in the present experiment.

A parallel can be made between the results of the present experiment and previous investigations into the MAE (Boutet et al., 1996; Chaudhuri, 1990;

Lankheet & Verstraten, 1995; Zhou & Chen, 1994). In particular, the finding that attending to motion during adaptation increases the MAE relative to attending to static stimulus parameters (Boutet et al., 1996) may be considered analogous to the increase in temporal recalibration when observers attend to the temporal relationship between adapting stimuli reported here. The data presented here suggest that increased aftereffect size by attending to the adapting parameter may be a general feature of sensory processing.

In summary, the present experiment shows that observers' attention can dramatically modulate TOJs made subsequent to asynchrony adaptation. In particular, attending to the temporal order of the adapting stimuli appears to produce significantly greater effects on subsequent TOJs compared to when attention is focussed on non-temporal aspects of the task. In addition, the results recorded here provide a potentially useful framework with which to explain inter-study differences in the magnitude of asynchrony adaptation effects.

# Chapter 7

## 7.1 Effects of sensory modality and causality on sensorimotor temporal order judgments

### 7.1.1 Introduction

In the course of everyday life, humans usually act with the intention to cause an effect. Often, this effect is experienced as a sensory event. For example, at the end of a concert an individual may decide to show their appreciation for the performance by clapping their hands in applause. In performing such a routine action, it can be seen that the sight of the hands making contact, the resultant clapping sound, and the induced tactile sensation all appear to coincide in time – in other words, the physically simultaneous events are correctly perceived as occurring simultaneously. This occurs despite differences in sensory latency between the modalities (Sections 1.1.1-1.1.3). Similarly, when using a computer there is typically no perceived delay between the physical press of a mouse button (voluntary action) and the corresponding on-screen response (sensory event). This is despite the fact that use of the mouse does not instantly translate to a response on the computer screen; a small hardware-dependent delay between use of the mouse and corresponding on-screen response is unavoidable. This delay is frequently in the order of 35msec (Cunningham et al., 2001; Stetson et al., 2006). It can therefore be seen that the overwhelmingly veridical percept of our actions and their effects when using a computer is actually erroneous. This contrasts with the correct, veridical percept of our voluntary actions and their effects (e.g., clapping the hands together) that predominates in a natural environment.

Despite this hardware-dependent delay between a motor action and its on-screen sensory consequence, our experience of using a computer is consistent with there being no perceived lag between a press of a mouse button and the on-screen response. In fact, previous work suggests that when a sensory event (visual or auditory) follows a voluntary action (pressing a button on a mouse) within 0-10msec, observers perceive the action as being 'too fast' in relation to the event, and that the subjectively 'most comfortable' interval between buttonpress and feedback is as long as 114-170msec, dependent on sensory modality and experimental paradigm (Nittono, Shimizu, & Hori, 2004). However, the subjective sensation of 'comfort' in this context (measured via a rating scale) is not necessarily comparable to conventional measures of perceived simultaneity or temporal order.

The fact that a unified percept of a voluntary action and its effect predominates under most circumstances in spite of the potential obstacles discussed above suggests the existence of a perceptual mechanism(s) effecting temporal realignment of voluntary actions and/or events. Such a hypothesis is supported by previous work. For example, the process of intentional binding has been shown to cause motor actions and their sensory consequences to shift towards each other in perceived time (as discussed in

Section 2.1.6). The available evidence strongly suggests that the intention to cause an effect is critical for this effect to occur (Haggard et al., 2002; Tsakiris & Haggard, 2003).

As well as intention, it has been shown that recent experience can also influence the temporal relationship between an action and its sensory consequence. Cunningham and colleagues measured observers' performance on a computer game, with on-screen movements controlled by a mouse (Cunningham et al., 2001); a snapshot showing the task performed by observers can be seen in Figure 7.1. After initially performing the task with a 35msec delay (the smallest possible on their computer) between movement of the mouse and movement on the computer monitor (a delay which was unnoticeable to observers), a marked reduction in performance was observed when the 35msec delay between action and effect was increased to 235msec. However, performance rapidly recovered to baseline levels after a period of practice.

Of particular interest are the unsolicited comments made by several observers and recorded in the work of Cunningham et al (2001). It was reported that although the 235msec delay between action and effect was initially very obvious to observers, at the end of data collection the delay was no longer perceived - in other words, action and effect appeared to temporally coincide, despite being physically separated by 235msec. When the 235msec delay was changed back to 35msec, some observers reported an apparent reversal of perceived temporal order: the on-screen object appeared to move *before* the mouse movement which caused it (Cunningham et al., 2001). Taken together, these comments strongly suggest that humans can adapt to a physical delay between a motor action and its sensory consequence in a way that the delay becomes imperceptible, and that an illusory reversal of perceived temporal order may represent an aftereffect of this phenomenon.

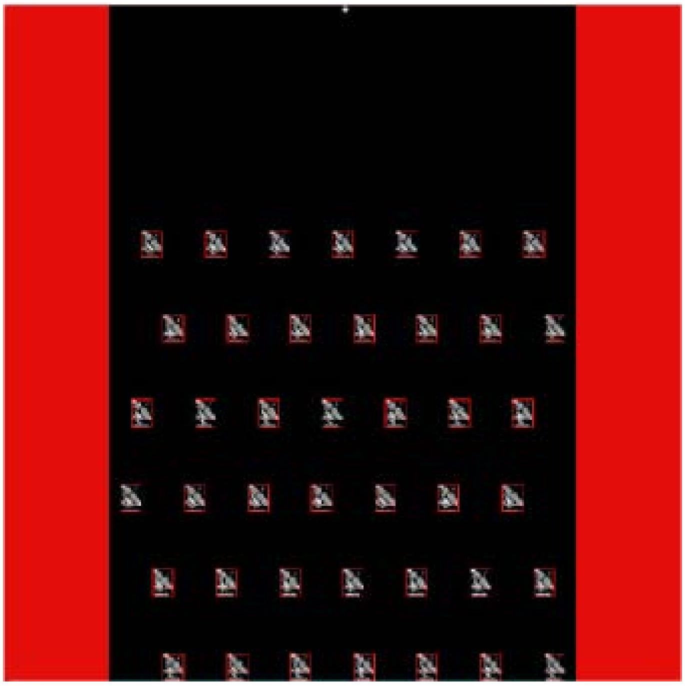


**Figure 7.1.** Snapshot of the task employed by Cunningham et al. (2001) to investigate the ability of humans to adapt to a new temporal relationship between their voluntary action and its consequence. Observers controlled a white aeroplane (just visible at the top of the snapshot) using a mouse to negotiate an obstacle field (seen in the lower half of the snapshot), which advanced at a constant speed towards the aeroplane. Performance was measured by the fastest speed at which observers could consistently negotiate the obstacle field successfully (Cunningham et al., 2001).

More recent work (Stetson et al., 2006) has empirically investigated this illusory reversal of perceived temporal order of action and effect. The overall pattern of results confirmed the anecdotal observations reported by Cunningham et al. (2001). This study found that following adaptation to a fixed delay of 135msec between keypress and flash, a flash that appeared within an average 64msec temporal window following a keypress was perceived as occurring *before* the keypress which caused it (Stetson et al., 2006). An investigation of the temporal tuning of the illusion found an inverse relationship between the magnitude of the illusion and increasing delay between voluntary action (keypress) and sensory consequence (flash); the effect appeared to be greatest with a 135msec delay, reduced slightly with a 285msec delay, and statistically insignificant or absent with a 535 or 1035msec delay (Stetson et al., 2006). Such temporal tuning is consistent with speculation that longer delays between voluntary action and effect modulate a sense of causality on the part of the observer; specifically, an observer is progressively less likely to perceive a sensory event as caused by their action with increasing delay between action and effect (Eagleman & Holcombe, 2002; Liddle & Jackson, 2006). Similar temporal tuning has been demonstrated in intentional binding, where the magnitude of the recorded temporal shifts declines with increasing delay between action and effect (Haggard et al., 2002).

The illusory reversal of perceived temporal order discussed above has been identified using visual test stimuli (Cunningham et al., 2001; Stetson et al., 2006). However, our motor actions frequently cause feedback in other (or multiple) sensory domains; as yet, whether the illusion is also apparent in these other sensory domains or is confined to the visuo-motor condition is unknown. Visual stimuli such as those employed by Stetson et al. (2006) and Cunningham et al. (2001) appear more likely to be subject to perceptual distortions with low-level neural loci than auditory or tactile stimuli. For example, perceived visual duration is compressed in a spatially-specific manner following adaptation to drifting gratings (Burr et al., 2007; Johnston et al., 2006). Similar reductions of perceived duration can be observed by simply reducing the visibility or increasing the spatial frequency of visual stimuli (Terao et al., 2008). Finally, compressions of both time and space around the time of saccadic eye movements have been shown to occur for visual, but not auditory, stimuli (Morrone et al., 2005). Given these distributed, low-level timing mechanisms, it is possible that the temporal illusion recorded by Stetson et al. (2006) and Cunningham et al. (2001) is observed only when observers' actions produce visual sensory consequences. This would suggest that the illusion is mediated by a low-level, modality-specific perceptual mechanism.

However, the possibility of Stetson et al.'s effects being confined to the visuo-motor domain is rendered less likely by the recent finding that intentional binding effects appear to occur between motor actions and their visual, auditory and tactile sensory consequences (Engbert et al., 2008). Thus, perceived sensorimotor timing appears flexible in visuo-motor, auditory-motor and tactile-motor contexts. In addition, purely sensory (as opposed to sensorimotor) temporal recalibration occurs in a manner consistent with a single, supramodal perceptual mechanism mediating adaptive temporal shifts in the audiovisual (AV), audiotactile (AT) and visuotactile (VT) sensory pairings (Sections 5.2.4 and 5.3.3). Finally, temporal perceptual learning effects appear to be specific to the trained temporal interval but transfer readily between sensory modalities (Nagarajan et al., 1998) and between perceptual and motor tasks (Meegan et al., 2000; Pesavento & Schlag, 2006).

These findings are more consistent with late-stage timing mechanisms operating independently of sensory modality or task. If the illusion reported by Stetson et al. (2006) and Cunningham et al. (2001) was mediated by such a mechanism, it would be expected that the illusion would also be observed in the auditory-motor and tactile-motor sensorimotor pairings. To investigate this possibility, a task to investigate the illusory reversal of perceived temporal order of action and effect described previously (Cunningham et al., 2001; Stetson et al., 2006) was devised.

Previous work on the perceived time of voluntary actions and/or their effects has overwhelmingly used the paradigm devised by Libet and colleagues (Libet et al., 1983). This paradigm is discussed in detail in Section 3.3.4. Due to the unresolved controversy surrounding Libet’s paradigm, and the finding that errors associated with timing judgments made using this method vary between the sensory modalities (Danquah et al., 2008), this method was deemed unsuitable for use in the present study.

As the proposed experiments were concerned with an illusory reversal of perceived temporal order of action and effect, rather than perceived timing of actions/effects or intentional binding *per se*, the methodology of the two papers which demonstrated such an effect was considered (Cunningham et al., 2001; Stetson et al., 2006). The work of Cunningham and colleagues only anecdotally suggested the existence of such an illusion, and it is difficult to see how their experimental paradigm (measuring performance on a computer game) could be adapted to rigorously quantify such effects.

Stetson and colleagues (2006), however, used a paradigm that yielded quantitative measurements of the illusion in question. Observers made a voluntary action (keypress) in response to a cue, and were presented with visual feedback (the sensory consequence of the action). The visual feedback was presented at a range of delays, but 60% of trials were with a delay of either 35 ('baseline' condition) or 135msec ('injected delay' condition); a 35msec delay was the minimum possible with the computer used in the experiment. The timing of the remaining 40% of visual events was controlled by tracking observers' reaction time to the cue and presenting the flash before or after the predicted time of the keypress. Observers had to respond whether the flash appeared to occur before or after the keypress (i.e., make a sensorimotor temporal order judgment (TOJ)). Observers' responses were used to plot a psychometric function from which the *physical* temporal offset between motor action and sensory event where the two appear *perceptually* simultaneous (i.e., sensorimotor PSS) and JND values were derived.

From consideration of the method of Stetson et al. (2006), it is apparent that a significant proportion of the test presentations occurred physically *before* the keypress which 'caused' them. This impressive feat was managed through ongoing measurement of the time elapsed between presentation of the initial (unspecified) cue and the keypress, generating a mean reaction time (RT) which was used to estimate the point at which the keypress was likely to occur. However, it should be noted that RT measurements have a degree of inherent variability, in common with all psychophysical measurements. Given that on 40% of trials the timing of visual probe presentation was dependent on this reaction time, it is obvious that there is a degree of uncertainty regarding the precise temporal location of 40% of the test presentations. Although this potential confound is unlikely to affect the broad pattern of results, it was felt that such a potential source of unnecessary variance was undesirable in the proposed study. Additionally, given the significant role of intention and causality in the perceived timing of motor actions and sensory events (Haggard et al., 2002; Tsakiris & Haggard, 2003), it may be considered undesirable to present sensory events physically *before* motor actions as the two events may not then be causally linked. Finally, motor actions performed in response to external events ('stimulus-based'), such as those in response to a sensory cue as in the paradigm of Stetson et al. (2006), are thought to employ a different neural substrate than those executed without such an external cue and with the aim of causing an external event ('intention-based') (Bueti et al., 2008c; Keller et al., 2006; Waszak et al., 2005).

Instead, it was decided to use a novel TOJ paradigm which eliminated these potential confounds by presenting *all* stimuli after the observers voluntary action, thereby allowing precise control of stimulus timing and ensuring that the sensory event could be interpreted as a consequence of observers' actions. Experimentation with the existing computer setup revealed that the minimum delay which was reliably reproducible between a buttonpress and a LED flash, auditory click or solenoid tap was 25msec. Therefore, motor action and sensory event would always be separated by a minimum of 25msec throughout the proposed experiment. The previous descriptions of the illusion of interest (Cunningham et al., 2001; Stetson et al., 2006) were compelling enough such that, irrespective of any previous potential methodological confounds, it was likely that such an illusion would also be recorded with the current experimental procedure despite the fact that at no point were stimuli presented physically prior to motor actions. Therefore, it seemed likely that the presentation of all test stimuli following the keypress would still result in an acceptable fit of the psychometric function to the data points. If, on the other hand, an illusion was not observed with this experimental paradigm, the percentage of ‘event appeared to precede action’ responses at the minimum-possible delay (25msec) would fail to reach chance performance.

In summary, the intention of the proposed experiment was to use a sensorimotor TOJ paradigm to ascertain whether a reversal of perceived temporal order of action and effect could be elicited in the auditory-motor and tactile-motor domains. It was hoped that comparison of these results with those in the visuo-motor domain would offer information as to the nature of the underlying sensorimotor timing mechanism(s).

## 7.2 Investigation of illusory reversal of perceived sensorimotor temporal order in the visuo-motor, auditory-motor and tactile-motor sensorimotor pairings

### 7.2.1 Observers

Observers were author JVMH, DW and JH and trained naïve observers AD and IP. All had extensive experience of sensory TOJs, and undertook a brief training period (~1hour) using a sensorimotor TOJ task prior to data collection.

### 7.2.2 Methods and stimuli

Visual, auditory and tactile stimuli were identical to those previously described (Section 4.2.2). Presentation of the stimuli was controlled by custom-written software run in MatLab (Mathworks, U.S.A.) on a desktop PC. The relative timings of all stimuli were verified by simultaneous capture on a multiple trace oscilloscope.

### 7.2.3 Procedures

The experiment was conducted in a darkened room. Observers wore tight, pinna-enclosing headphones throughout each experimental run. This rendered the noise of the mousepress inaudible to observers, therefore preventing observers using this noise as an additional cue to judge the time of their action. During the experiment, observers pressed a mouse button at a pace of their own choosing in order to ensure their actions were intention-based in nature (Bueti et al., 2008c; Keller et al., 2006; Waszak et al., 2005).

After each of the first four mousepresses, a stimulus was presented (e.g., LED flash) at a constant delay of 100msec (‘adapting’ presentations). This value was chosen due to previous work suggesting that the temporal illusion of interest should be apparent when adapting to such a delay (Stetson et al., 2006). After the fifth mousepress, the same stimulus was presented (‘test’ presentation) with a delay of 25, 50, 75, 100, or 125msec, which varied randomly within a method of constant stimuli. Observers were required to judge whether the fifth stimulus presentation appeared *before* or *after* the fifth buttonpress, and make an unspeeded response via a computer keyboard. The modality of the adapting and test stimuli remained constant throughout each experimental run. A schematic explaining the experimental paradigm is shown in Figure 7.2. Each of the five delays was tested 10 times within an experimental run. Observers completed five runs for each of the three sensorimotor pairings tested, making a total of 750 test presentations per observer (10 presentations* 5 runs* 5 test delays * 3 sensorimotor pairings).

For each observer and condition, the percentage of ‘test stimulus before buttonpress’ responses was plotted as a function of test delay and fitted with a logistic function identical to that described previously (Section 4.3.3). In this way, PSS values were obtained for all observers in all of the conditions. In the present work, PSS represents the test delay at which observers’ voluntary actions were perceived as occurring simultaneously with a subsequent sensory event. The data were analysed using KaleidaGraph v4.0 (Synergy Software, U.S.A.).

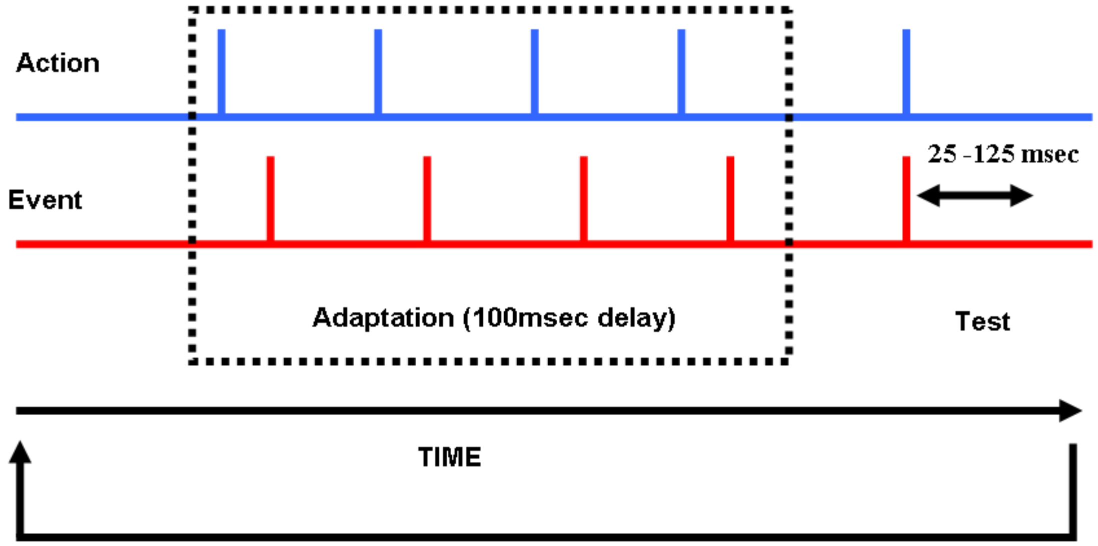


**Figure 7.2.** Schematic showing the paradigm used in the current experiment. Observers pressed a button four times at a pace of their own choosing ('Adaptation'), with each press causing a sensory event at a fixed delay. The fifth buttonpress caused the same event, but with a delay of 25, 50, 75, 100 or 125msec ('Test'). Observers were required to make a forced-choice response as to whether they perceived the fifth event as occurring *before* or *after* the buttonpress.

### 7.2.4 Results

Figure 7.3 shows psychometric functions for observer JVMH for the visuo-motor, auditory-motor and tactile-motor sensorimotor pairings. The percentage of trials where the observer perceived a sensory event as preceding a voluntary action is plotted on the y-axis. Since the voluntary action *always* caused the sensory event, such a response is suggestive of an illusory reversal of perceived temporal order of action and event. The test delay corresponding to 50% 'event before action' responses is the sensorimotor PSS. As can be seen from the figure, this observer perceives a visual event occurring 48msec after a voluntary action as being simultaneous with the action. Similarly, an auditory event occurring 65msec after, and a tactile event occurring 49msec after the same voluntary action, are perceived as simultaneous to the action.

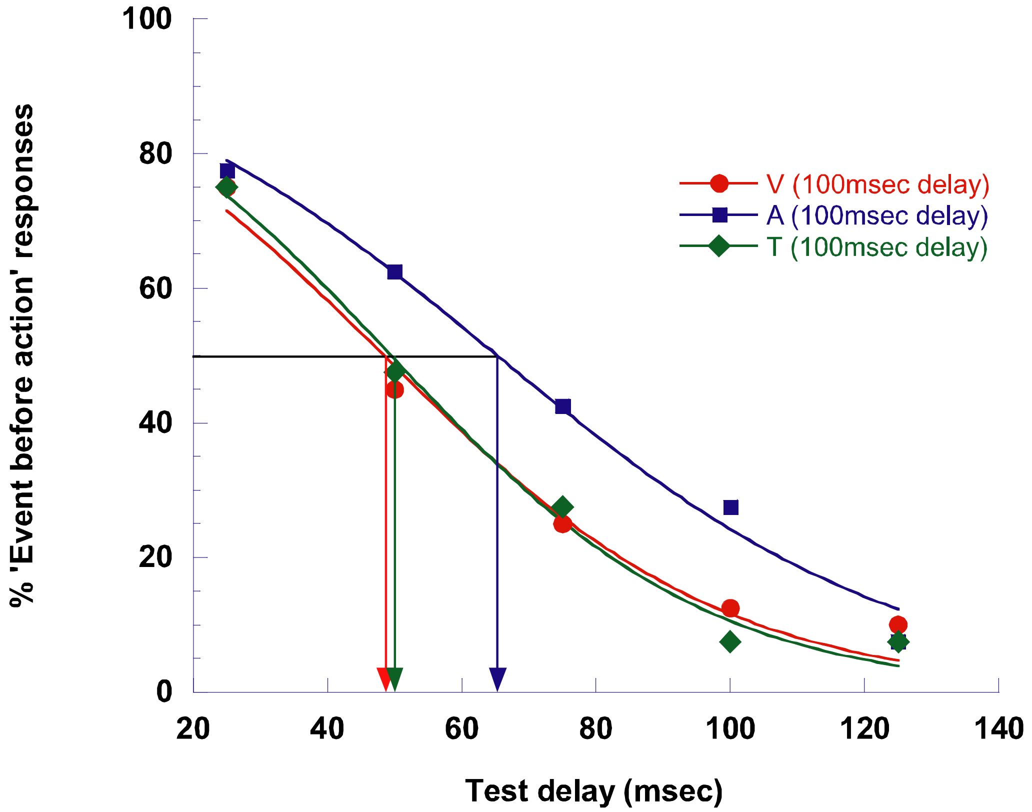


**Figure 7.3.** Psychometric functions derived from data collected by observer JVMH in the visuo-motor (V, red circles), auditory-motor (A, blue squares), and tactile-motor (T, green diamonds) sensorimotor pairings. The mid-points of the functions represent the point at which the observer perceives the motor action and sensory event as occurring simultaneously – the Point of Subjective Simultaneity (PSS). The test delay corresponding to the PSS is indicated for each modality by the arrows on the x-axis.

PSS values averaged across all five observers are presented as Figure 7.4. The figure shows that the pattern of results for author JVMH, as seen in Figure 7.3, is reproducible across observers, and that on average observers perceive a voluntary motor action followed by a sensory event 61 (V), 67 (A) or 58 (T) msec later as occurring simultaneously. A one-way repeated measures ANOVA showed that within observers, PSS did not differ significantly in the three pairings ($F_{2,\,8} = 1.17$, $p > 0.05$).

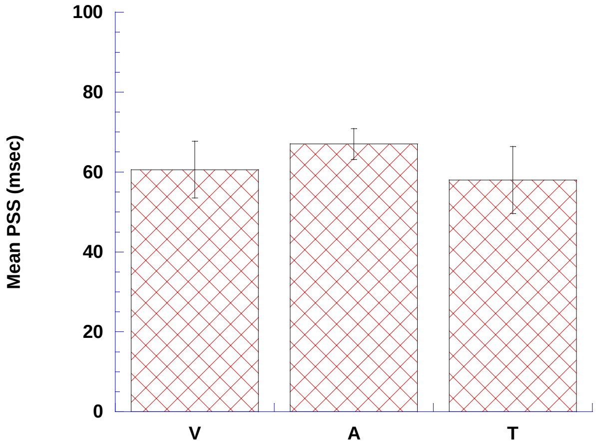


**Figure 7.4.** Averaged PSS values for all five observers in the visuo-motor (V), auditory-motor (A) and tactile-motor (T) sensorimotor pairings following adaptation to a 100msec delay between action and sensory event. Error bars represent one standard error of the mean (SEM) either side of the parameter values

JND values averaged across all five observers are presented as Figure 7.5. It can be seen that mean JNDs are approximately equal in all three sensorimotor pairings. This was confirmed by a one-way repeated measures ANOVA ($F_{2,\,8} = 0.326$, $p > 0.05$).

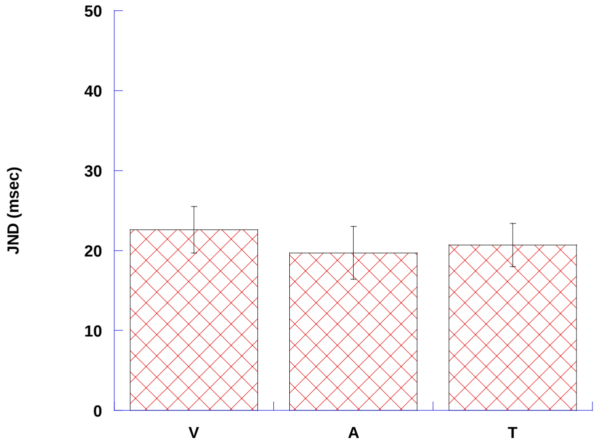


**Figure 7.5.** Averaged JND values for all five observers in the visuo-motor (V), auditory-motor (A) and tactile-motor (T) sensorimotor pairings following adaptation to a 100msec delay between action and event. Error bars represent one SEM either side of the parameter values.

### 7.2.5 Discussion

The results of the present experiment clearly demonstrate, for the first time, that the temporal illusion previously described in the visuo-motor pairing (Cunningham et al., 2001; Stetson et al., 2006) has parallels in the auditory-motor and tactile-motor pairings.  The replication of previous results in the visuo-motor pairing with the use of a different paradigm suggests that the observed illusion is robust in nature.

Task difficulty, as evidenced by the average JND values presented in Figure 7.5, is comparable across the three sensorimotor pairings.  Examination of average PSS values (Figure 7.4) shows that the magnitude of the illusion is also comparable across the three sensorimotor pairings.  This is suggestive of either a single common neural mechanism mediating the observed effects, or distinct individual mechanisms which exhibit similar characteristics when observers adapt to a delay between action and effect of 100msec.  In order to investigate these possibilities further, it was decided to ascertain the temporal tuning of the illusion in all three sensorimotor pairings.  Similar temporal tuning in all three pairings would provide additional support for a single neural mechanism mediating the observed effects.  Conversely, inter-pairing differences in the temporal tuning of the illusion would be more consistent with separate mechanisms.

In the context of the present temporal illusion, temporal tuning is highly likely to reflect the influence of intention. If the attribution of intentionality (the intention to cause an effect) is essential for the observed temporal realignment to take place, then it is reasonable to assume that the magnitude of temporal recalibration will decline with adaptation to an increasing delay between action and effect. This assumption is based on the fact that an observer is less likely to perceive a sensory event as being caused by their action with an increasing delay between action and event (Eagleman & Holcombe, 2002; Franck et al., 2001; Liddle & Jackson, 2006). If so, it is likely that should observers adapt to a delay between motor action and sensory event larger than that within which observers attribute causality to a sensory event, no recalibration effects/temporal illusion would be observed. This would be shown by a PSS value of approximately zero.

With these factors in mind, an experiment to assess the temporal tuning of the illusion in question was devised.

## 7.3 Temporal tuning of illusory reversal of perceived sensorimotor temporal order

### 7.3.1 Methods and stimuli

Observers and stimuli used were the same as from the previous experiment (Sections 7.2.1 and 7.2.2). As before, AD and IP were naïve as to the purpose of the experiment.

### 7.3.2 Procedures

The experimental procedure was identical to that in the previous experiment (Section 7.2.3), except that the adapting delay was 50, 200, 400 or 800msec. This adapting delay remained constant throughout each experimental run. The twelve experimental conditions (3 sensorimotor pairings * 4 adapting delays) were performed in a random order, with the condition changing after each run to minimise any possible effects of practise on analysis of JND. As before, observers were required to make forced-choice judgements as to whether the fifth (test) stimulus presentation appeared *before* or *after* the fifth buttonpress, and respond in an unspeeded manner via the computer keyboard. Each of the five test delays (25, 50, 75, 100 and 125msec) was tested ten times within an experimental run. Each observer completed five runs in each of the 12 experimental conditions, making a total of 3000 test presentations per observer (10 presentations * 5 runs * 5 test delays * 4 adapting delays * 3 sensorimotor pairings).

The data for each experimental condition were plotted and analysed in an identical manner to that described previously (Section 7.2.3).

## 7.3.3 Results

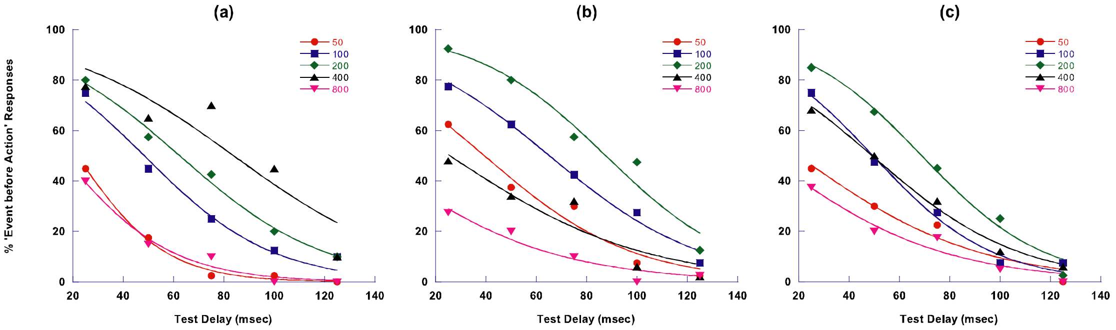


**Figure 7.6 (a-c).** Psychometric functions for observer JVMH derived from sensorimotor TOJ data following adaptation to various delays between motor action and sensory event in the visuo-motor (a), auditory-motor (b) and tactile-motor (c) pairings. For each sensorimotor pairing, the individual functions relate to an adapting delay of 50 (red circles), 100 (blue squares), 200 (green diamonds), 400 (black triangles) or 800msec (pink inverted triangles). The curves relating to an adapting delay of 100msec are identical to those in Figure 7.3.

Psychometric functions for observer JVMH can be seen in Figure 7.6(a-c) for the visuo-motor (a), auditory-motor (b) and tactile-motor (c) sensorimotor pairings. For completeness, this observer’s functions from Section 7.2.4 (with an adapting delay of 100msec) are included on the figure. Inspection of the graphs reveals a clear separation between the functions corresponding to different adapting delays. Consider Figure 7.6(b), red curve (which relates to an adapting delay of 50msec between motor action and auditory event): it can be seen that the blue and green curves (100 and 200msec adapting delays, respectively) are displaced rightward, indicating an increased PSS (and therefore, a larger magnitude of the illusion). Conversely, the black and pink curves (400 and 800msec adapting delays, respectively) are displaced leftward relative to the green 200msec curve (indicating a decreased PSS/reduced magnitude of illusion). The manner of the separation of the functions suggests that the illusory reversal of perceived temporal order of action and effect is temporally tuned in all three sensorimotor pairings.

PSS values were averaged across all observers and plotted as a function of adapting delay for each sensorimotor pairing. The resultant averaged PSS tuning curves for the visual (red circles/curve), auditory (blue squares/curve) and tactile (green diamonds/curve) modalities are shown in Figure 7.7. Each data point represents the average PSS for five observers. It is clear that the illusion is temporally tuned in each of the three sensorimotor pairings, and a strong similarity in the shape of the tuning functions is evident. The curve fit is described in Section 7.5.

A repeated-measures ANOVA on the PSS values found an expected and highly-significant effect of adapting delay ($F_{4, 16} = 37.1$, $p < 0.001$), confirming the temporal tuning of the illusion in each sensorimotor pairing apparent from inspection of Figure 7.7. The effect of sensory modality also reached significance ($F_{2, 8} = 5.57$, $p < 0.05$), reflecting the slightly lower magnitude of the illusion in the tactile-motor pairing. These factors did not interact ($p > 0.05$), confirming the similarity in the effects of adapting delay on the illusion in the three sensorimotor pairings. Finally, a repeated-measures ANOVA revealed that JND values did not differ significantly between the three sensorimotor pairings ($F_{2, 8} = 1.23$, $p > 0.05$).

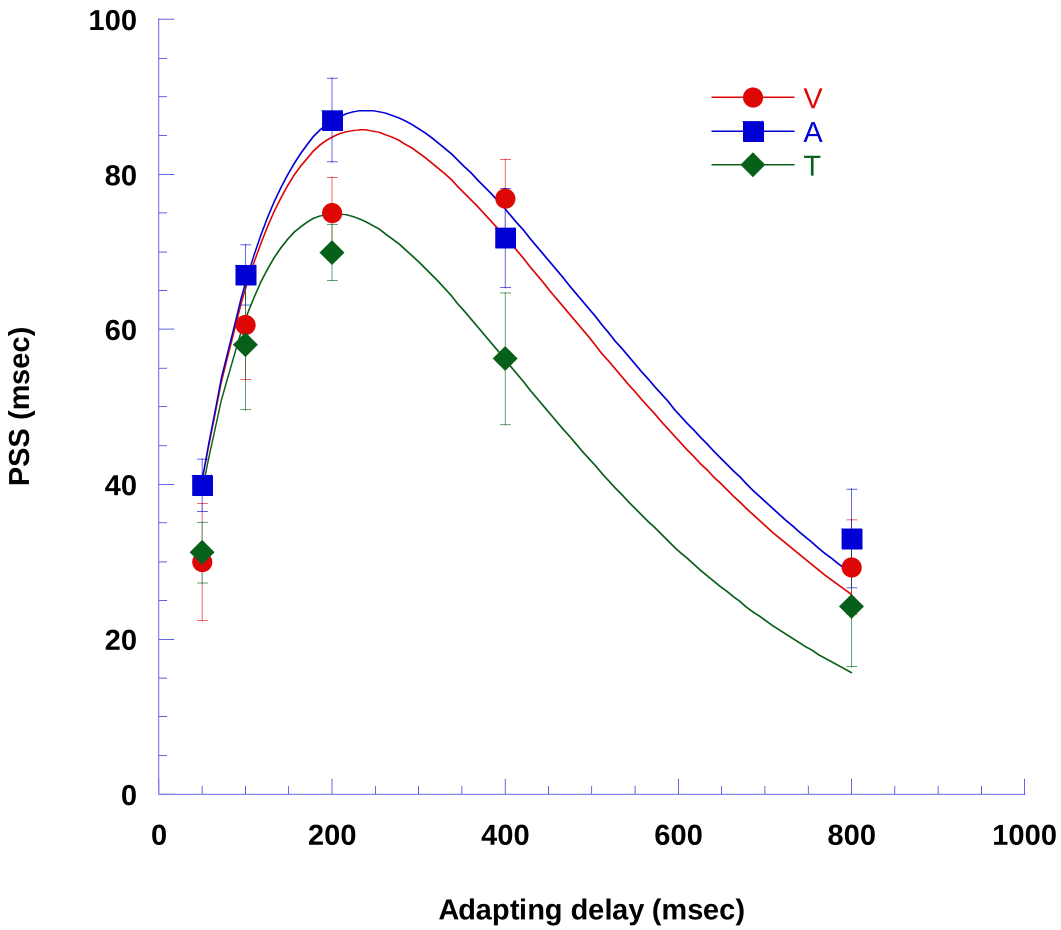


**Figure 7.7.** Diagram showing the temporal tuning of an illusory reversal of perceived temporal order of action and a sensory event in the visual (red circles), auditory (blue squares) and tactile (green diamonds) sensorimotor pairings. Each data point represents the mean PSS (n = 5), or the average physical delay at which action and effect are perceived as simultaneous, at each of five adapting delays: 50, 100, 200, 400 and 800msec. The curve fit is described in Section 7.5.1. Error bars represent one SEM either side of the parameter values.

### 7.3.4 Discussion

Figure 7.7 and the analysis of PSS values clearly show that the illusion demonstrated in Section 7.2.4 and elsewhere (Cunningham et al., 2001;

Stetson et al., 2006) is temporally tuned in all three sensorimotor pairings tested. The three tuning functions appear very similar in shape (confirmed by the lack of interaction between adapting delay and test stimulus modality), indicating that modulating the causal link between action and event (via varying the adapting delay) changes the illusion in a comparable manner in all three pairings. This finding, combined with the close correspondence in JND values, is consistent with the observed temporal illusion being mediated by a single perceptual mechanism serving the three sensorimotor pairings tested. An obvious parallel can be made with purely sensory TOJs (Section 5.4.3).

The temporal tuning curves also strongly argue against the potential criticism that the illusion recorded here is simply an artefact of response bias and the experimental paradigm employed. As all sensory events were presented physically after the causative buttonpress, the results presented in Section 7.2.4 could also have been obtained had observers unconsciously adopted a strategy of balancing their sensorimotor TOJs (e.g. 50% ‘buttonpress first’, 50% ‘sensory event first’), thereby forcing the PSS to be centred upon the mid-point of the range of temporal offsets between the stimuli. If this were the case, however, the tuning functions would approximate horizontal lines (i.e., PSS values would be approximately identical for each adapting delay); Figure 7.8 shows that this is definitively not the case.

From the results of the present experiment, it appears that the adapting delay is crucial to the illusory reversal of perceived temporal order. The nature of sensory feedback appears less critical, with no significant effects of modality on JND and little effect of modality on PSS values. If sensory modality is a genuinely unimportant variable, it is likely that adaptation temporally recalibrates the relationship between a motor action and a subsequent sensory event in *any* modality. Should this be the case, a change in the modality of sensory feedback between the adapt and test phases would be expected to have little or no effect on the illusion reported here. Such a finding would provide further support for the existence of a single, supramodal perceptual mechanism mediating recalibration of the perceived temporal relationship between action and sensory feedback in all of the tested sensorimotor pairings. If, conversely, no temporal recalibration was observed when altering the modality of the sensory event between adapt and test phases, this would suggest distinct but functionally similar mechanisms mediating the illusion in the visuo-motor, auditory-motor and tactile-motor pairings.

To test these possible outcomes, it was decided to systematically measure sensorimotor PSS following adaptation to a delay between an action and subsequent event in one modality, whilst observers judged the time of the action relative to a sensory event presented in a different modality.

## 7.4 Transfer across modalities of illusory reversal of perceived sensorimotor temporal order

### 7.4.1 Methods and stimuli

Observers and stimuli were the same as in the previous experiments (Sections 7.2.1, 7.2.2 and 7.3.1), with the addition of two more trained naïve observers DL and MB; these observers were experienced at sensory TOJs and were trained in sensorimotor TOJs in the manner described previously (Section 7.2.1).

### 7.4.2 Procedures

In contrast to the previous two experiments (Sections 7.2.3 and 7.3.2), the adapting and test stimulus modalities differed. A total of six crossmodal conditions (adapt V test A, adapt V test T, adapt A test V, adapt A test T, adapt T test V, and adapt T test A) were tested, with a fixed adapting delay of 200msec in all conditions. This value was chosen on the basis of the results shown in Figure 7.7, which indicated a robust reversal of perceived temporal order after adaptation to this delay between voluntary action and sensory event. Therefore, should the illusion still exist when changing modality between adapt and test phases, it should be quantifiable when using a 200msec adapting delay. In all other respects, the procedure was identical to that employed in Section 7.2.3. Each observer completed five runs in each of the six crossmodal conditions, making a total of 1500 test presentations per observer (5 test delays * 10 presentations * 5 experimental runs * 6 conditions). The data for each experimental condition were plotted and analysed in an identical manner to that described previously (Section 7.2.3).

### 7.4.3 Results

Psychometric functions for observer JVMH are shown in Figure 7.8(a-c, corresponding to visual, auditory and tactile test events respectively). On each graph, the green data points and functions (representing sensorimotor TOJs made when adapt and test stimuli were presented in the same modality) are derived from the data of the previous experiment (Section 7.3.3). Inspection of the functions indicates that the illusion still occurs when adapt and test stimuli are presented in different modalities. The functions appear to be slightly steeper when adapt and test stimuli were presented in the same modality (green diamonds/functions) than in separate modalities (red circles/functions and blue squares/functions), suggesting increased sensitivity to sensorimotor asynchrony when adapting and test stimuli are identical. This is particularly apparent from inspection of Figure 7.8(c), which relates to conditions where the test stimulus was always in the tactile modality.

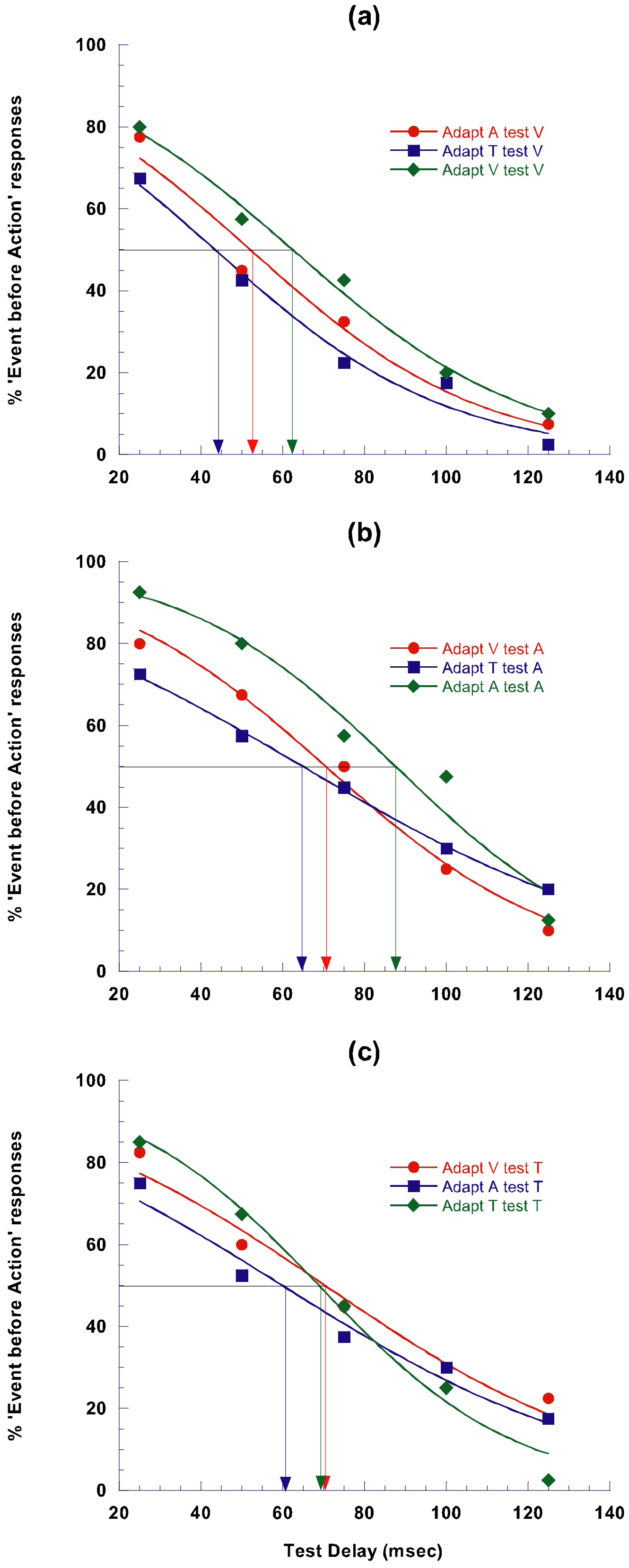


**Figure 7.8(a-c).** Psychometric functions derived from data collected by observer JVMH, plotted according to test stimulus modality: visual, auditory and tactile (a-c, respectively). The test delay corresponding to the mid-points of the functions represent the PSS. In all cases, the adapting delay between motor action and sensory event was 200msec. Green data points and functions are derived from the 200msec condition in the previous experiment (Section 7.3.3).

PSS values averaged across the seven observers are presented in Figure 7.9. Inspection of the Figure shows that mean PSS values are similar whether adapt and test stimuli are presented in the same or different modalities. PSS values for all seven observers were combined within a two-way, repeated-measures ANOVA, which revealed that the effect of test modality on PSS was significant ($F_{2, 12} = 7.41$, $p < 0.01$), but that the effect of the adapting modality was not significant ($F_{2, 12} = 1.61$, $p > 0.05$). Critically, however, there was no significant interaction between these two factors ($F_{4, 24} = 2.02$, $p > 0.05$). This indicates that the magnitude of the effect for a given test modality does not depend on the adapting modality. In other words, it does not matter whether the adapting modality is the same or different to that of the test modality – crossmodal effects are just as great as unimodal. The significant effect of test modality reflects the fact that the magnitude of the effect was slightly, but consistently, higher when the test stimulus was presented in the auditory modality (Figure 7.9(b)).

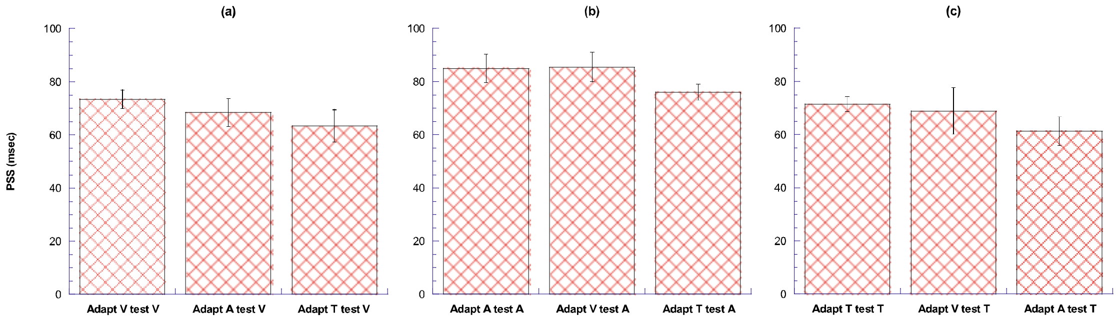


**Figure 7.9(a-c).** Average PSS values derived from seven observers when testing the visual (a), auditory (b) or tactile (c) modalities. The conditions where adapt and test stimuli were in the same modality are represented by the leftmost bar on each graph. Error bars represent one SEM either side of the parameter values.

JND values averaged from all seven observers are presented in Figure 7.10(a-c). The values for all observers were combined within a two-way, repeated-measures ANOVA. This revealed that both test modality ($F_{2,\ 12} = 0.389$, $p > 0.05$) and adapting modality ($F_{2,\ 12} = 0.769$, $p > 0.05$) had an insignificant effect on JND. However, the interaction between these factors was highly significant ($F_{4,\ 24} = 9.815$, $p < 0.001$), meaning that the JND for a given test modality depended critically on the adapting modality. This reflects the fact that, irrespective of the test modality itself, JNDs for conditions where adapt and test stimuli were presented in the same modality are lower than those in the crossmodal conditions (Figure 7.10(a-c), leftmost bars). There is clearly a cost to performance in having the modality of the sensory event change between adapt and test phases.

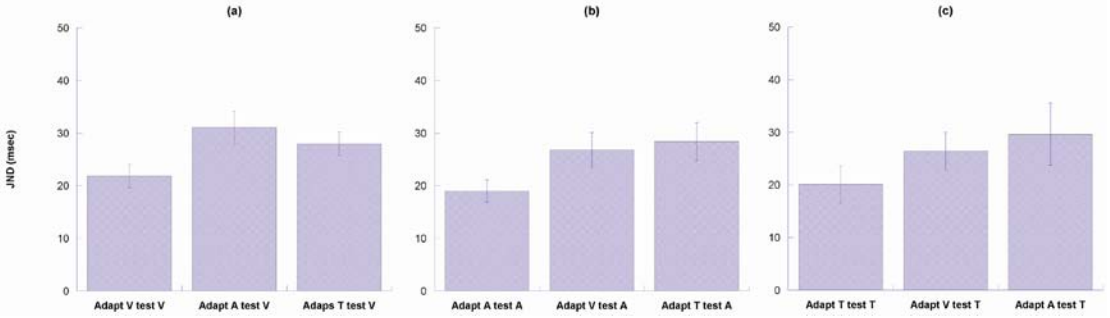


**Figure 7.10(a-c).** Average JND values derived from seven observers when testing the visuo-motor (a), auditory-motor (b) and tactile-motor (c) pairings. The conditions where adapt and test stimuli were in the same modality are represented by the leftmost bar on each graph. Error bars represent one SEM either side of the parameter values.

### 7.4.4 Discussion

The results of the present experiment confirm that adaptation to a delay between voluntary action and a sensory event modulates the perceived temporal relationship between an action and subsequent events in the same or (crucially) a *different* sensory modality. The effect on perceived temporal relationship of action and effect is equivalent whether the adapting and test modalities are the same or different. In other words, adaptation to a fixed delay between action and event creates a temporal window following the action, during which an event occurring in *any* of the three modalities tested appears to precede the action which caused it. This is the first empirical demonstration and quantification of such effects.

The fact that the illusion does not depend on whether the modality of the adapting stimuli is the same as that of the test stimulus, combined with indistinguishable JNDs in the three conditions, suggests that a single central, supramodal perceptual mechanism mediates the illusion in all three sensorimotor pairings. Such a mechanism is likely to operate at a relatively late stage of the neural processing hierarchy, beyond modality-specific cortical areas. A range of evidence exists to support this hypothesis.

Firstly, the temporal illusion in question has previously been shown to be independent of the low-level characteristics of the visual stimuli used; specifically, observers adapted to a delay between a voluntary action and a light which varied randomly between three different colours, whilst judging the timing of a light of a fourth different colour (Stetson et al., 2006). The illusion was still observed, suggesting that the illusion is unaffected by the chromaticity of the visual stimuli and thus not mediated by a low-level mechanism.

Similar evidence emerges from a study using a slightly different task – synchronisation of a motor action to a sensory event (Pesavento & Schlag, 2006). This study found that following adaptation to a delay between motor action and visual feedback, observers' performance in a different sensorimotor task reflected the recently-learned temporal relationship between action and event. In other words, the adaptation was not specific to the task used in adaptation. This result parallels the effects of purely sensory adaptation to asynchronous AV stimulus pairs, which has been shown to modulate observers' temporal perception measured with a range of different tasks and stimuli (Fujisaki et al., 2004).

This parallel between sensorimotor and sensory timing is consistent with other studies suggesting close links between temporal processing in the sensory and motor systems. For example, Ivry and Hazeltine found that the variability of motor timing was closely correlated with the variability of interval discrimination judgments over a range of durations (Ivry & Hazeltine, 1995). Additionally, Meegan and colleagues found that training in a (sensory) temporal interval discrimination judgment task significantly enhanced performance in a subsequent (motor) temporal interval reproduction task, and that the improvement was greatest when the temporal interval to be reproduced was the same as that used in the discrimination task. The authors suggested the existence of a plastic neural network shared by sensory and motor systems, possibly involving the cerebellum (Meegan et al., 2000). Further implication of the cerebellum as vital in both motor and sensory timing comes from fMRI studies (Bueti et al., 2008c; Schubotz, Friederici, & Yves von Cramon, 2000). However, note that Bueti et al. (2008c) found that additional cortical areas were recruited only during the motor timing task, whereas Schubotz et al. (2000) found almost complete overlap between motor and sensory timing activation patterns.

Two other features of the results are worthy of comment. Firstly, JND was found to be increased when the modality of the probe stimulus differed between adapt and test phases. This finding is likely to reflect the fact that temporal performance is compromised by rapidly shifting attention between sensory modalities (Spence et al., 2001a; Westheimer, 1999).

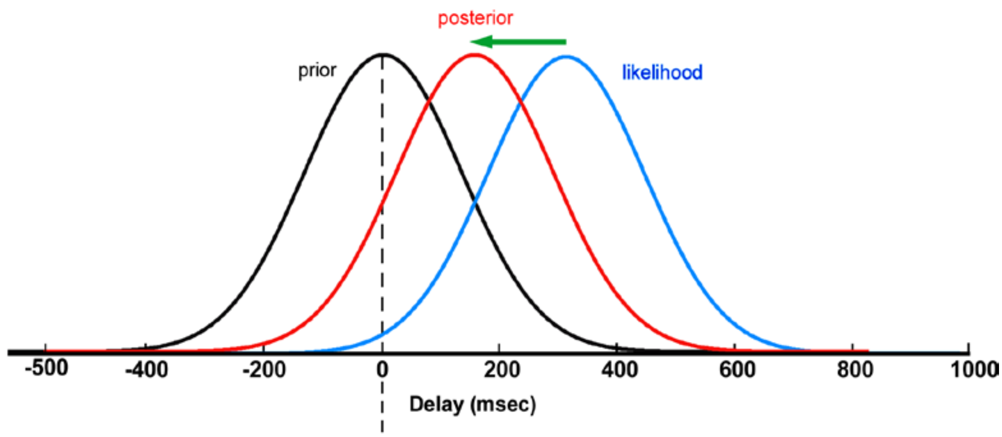
Additionally, the fact that PSS values (and thus the magnitude of the illusion) are higher when the test stimulus is auditory may be considered difficult to reconcile with the proposed single neural mechanism mediating the illusion. However, this may be a simple artefact of the properties of auditory temporal processing rather than suggestive of a separate mechanism serving the auditory-motor pairing. For example, filled and unfilled auditory intervals are perceived as longer than visual durations of the same objective length (Behar & Bevan, 1961 ; Goldstone & Goldfarb, 1963; N'Diaye et al., 2004; Walker & Scott, 1981; Wearden et al., 1998; Wearden et al., 2006) despite evidence that many of the same neural areas are involved in processing both visual and auditory duration information (Bueti et al., 2008a; N'Diaye et al., 2004). In other words, this difference in performance in the auditory-motor domain does not necessarily reflect a different perceptual mechanism. Given the temporal tuning and JND data, the overall weight of evidence is clearly suggestive of a single sensorimotor timing mechanism mediating the effects described here and elsewhere (Cunningham et al., 2001; Stetson et al., 2006).

## 7.5 A Bayesian explanation of illusory reversal of perceived temporal order of motor actions and their sensory consequences

### 7.5.1 Introduction

In order to make their sensorimotor TOJs in the present series of experiments, observers combined information from estimates of the timing of their motor action and subsequent sensory events into a percept of sensorimotor temporal order. An influential framework for understanding how observers combine such information in an efficient and coherent manner is maximum-likelihood estimation (MLE). In essence, MLE states that sensory information is combined in a manner that takes into account the reliability (inverse variance) of the individual cues (Ernst & Bulthoff, 2004). The perceptual outcome reflects the relative ‘weight’ accorded to the cues, with greater perceptual weight being given to the input(s) with the greater reliability. Bayesian decision theory (BDT) builds upon MLE by incorporating observers’ prior assumptions of the environment into the framework. Thus, the perceptual outcome (or ‘posterior’) of a Bayesian decision process reflects the strength of observers’ prior assumptions as well as the sensory information (‘likelihood’) itself.

**(a) Adaptation phase**



**(b) Test phase**

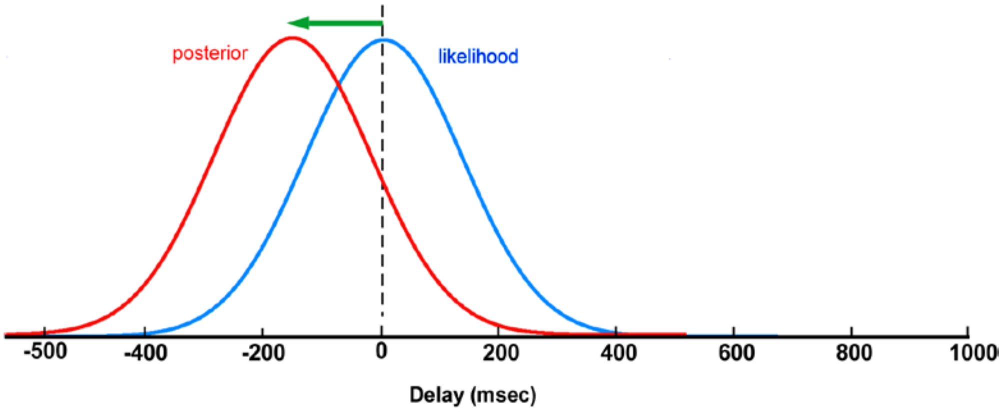


**Figure 7.11(a-b).** Schematic outlining a Bayesian framework for adaptive recalibration of sensorimotor temporal perception. In Figure 7.11(a), adaptation recalibrates the noisy sensorimotor temporal estimate (*likelihood function* – blue curve) via combination with an assumption of sensorimotor synchrony (*prior* – black curve). The perceptual outcome (*maximum a posteriori estimate’s* position relative to the *likelihood* - red curve relative to blue, respectively) is transferred over to the test phase (Figure 7.11(b)), where physically simultaneous sensorimotor test pairings are perceived as ‘event before action’. Green arrows show the magnitude of the temporal realignment.

In the absence of any prior knowledge about the statistical distribution of sensorimotor asynchronies in the external world, a Bayesian observer would employ MLE to ascertain the probability of different physical temporal orders, given a sensed temporal order (Mamassian, Landy, & Maloney, 2002). However, an absence of such prior knowledge represents an unlikely scenario (Eagleman, 2008; Eagleman & Holcombe, 2002). Throughout life, accumulated experience of the temporal relationship between voluntary motor actions and their sensory consequences will, by definition, be overwhelmingly dominated by physical simultaneity. If this experience builds up an *a priori* expectation of sensorimotor timing it will be centred on physical simultaneity (Figure 7.11(a) – black curve). Thus, we can expect a sensorimotor temporal order estimate (i.e., the maximum of the *likelihood function* – Figure 7.11(a) - blue curve) that does not match our expectation of simultaneity to be partially recalibrated to give *maximum a posteriori* (MAP) estimates closer to simultaneity (Figure 7.11(a) – red curve). The combination of the prior and likelihood allows MAP estimation via the creation of a Gaussian posterior distribution whose maximum ($x_{posterior}$) is defined by the weightings (w) assigned to the two component distributions as follows (after Miyazaki et al., 2006)[11]:

$$x_{posterior} = w_{prior} x_{prior} + w_{likelihood} x_{likelihood} \qquad \text{(Equation 1)}$$

---

[11] The derivation of Equation 1, and its relationship to a traditional Bayesian framework, is detailed in the Bayesian model described by Miyazaki et al. (2006) in their online Supplementary Methods section, available for download from the Nature website: www.nature.com/neuro/journal/v9/n7/suppinfo/nn1712_S1.html.

where $w_{likelihood} = 1 - w_{prior}$. Since the prior is centred on zero, $x_{prior} = 0$, hence

$$x_{posterior} = w_{likelihood} x_{likelihood} \qquad \text{(Equation 2)}$$

This represents the maximum of the posterior distribution relative to zero. The maximum of the posterior relative to that of the likelihood is therefore

$$\begin{aligned} & x_{likelihood} - w_{likelihood} x_{likelihood} \\ & = (1 - w_{likelihood}) x_{likelihood} \\ & = w_{prior} x_{likelihood} \end{aligned} \qquad \text{(Equation 3)}$$

The outcome of adapting to a given temporal delay between action and event is therefore to shift the perceived timing of the event to a position earlier than its true physical location in time, and Equation 3 quantifies the extent of this shift. During test trials (Figure 7.11(b)), when the delay between action and event is suddenly reduced, the relationship between the perceived and physical time (Figure 7.11 – green arrows) is carried over from the adapting phase, resulting in the misperception of event time having occurred prior to motor action. This is illustrated by the lateral separation between the red (MAP estimate) and blue (likelihood function) curves in Figure 7.11(b) (green arrow).

Given a constant weight of prior, Equation 3 predicts that the magnitude of the illusion should increase monotonically with delay between motor action and event during the adaptation phase. This is clearly not the case (Figure 7.7);

human observers typically show a rapid fall-off in their tendency to attribute sensory feedback as being a consequence of their motor actions as the temporal discrepancy between the two is increased (Asai & Tanno, 2007; Bays, Wolpert, & Flanagan, 2005; Blakemore, Frith, & Wolpert, 1999; Franck et al., 2001). The decision was therefore made to introduce a prior with a variable weight which decreases exponentially with increasing sensorimotor delay, and use this variable prior to perform a best fit to the empirical data for each of the three sensory conditions. Following Equation 3, the data of Figure 7.7 were fit with the following function:

$$PSS = \underbrace{e^{-\left(\frac{0.693 \times delay}{k}\right)}}_{w_{prior}} \times \underbrace{\mathbf{delay}}_{\mathbf{x_{likelihood}}} \qquad \text{(Equation 4)}$$

where the single free parameter, *k*, defines the rate of decline of $w_{prior}$ with increasing delay. It denotes the delay at which $w_{prior}$ falls to half its baseline value (i.e., $w_{prior} = w_{likelihood} = 0.5$). This value is similar for each of the three sensory modalities (161msec for visuo-motor, 166msec for auditory-motor and 141msec for tactile-motor). The model provides an excellent fit to the data sets for all three sensorimotor pairings, as shown in Figure 7.7. The quality of the fit is confirmed by $R^2$ values of 0.88 for visuo-motor, 0.98 for auditory-motor and 0.89 for tactile-motor curves.

### 7.5.2 Discussion

The experimental data discussed previously (Sections 7.2.4, 7.3.3 and 7.4.3) are well described by a Bayesian model in which an observer's prior experience leads them to impose a degree of perceptual synchrony between motor actions and their sensory consequences. One intuitively appealing feature of the model is that it balances the costs and benefits of recalibrating perception in response to an altered physical environment. Small sensorimotor temporal delays are treated as improbable, and the perceived sensorimotor timing of *all* potential sensory feedback (irrespective of sensory modality) is partially realigned, depending on the relative weightings of the synchrony prior and the likelihood. Increasing delays between actions and afferent sensory inputs are classified as exponentially more likely to have arisen from independent (i.e., external) causes, thus minimising the risk of erroneous adaptation.

Classic adaptation effects such as the motion aftereffect (MAE) (Mather et al., 1998) or the temporal recalibration reported in Sections 5.2.4, 5.3.3, 6.4.4 and 6.5.5, pose a problem for traditional Bayesian models. As recently pointed out (Clifford et al., 2007; Stocker & Simoncelli, 2006), perceptual realignment of pre-existing prior distributions would alter the $x_{prior}$ value in the direction of the $x_{likelihood}$ (a rightward shift of the black curve seen in Figure 7.11(a)), resulting in *attraction* rather than the archetypal *repulsion* effects observed throughout the adaptation literature. Interestingly, a recent study by Miyazaki and colleagues (Miyazaki et al., 2006) found that asynchronous taps to the spatially separated index fingers of left and right hands does indeed bring about attractive adaptation effects: observers imposed the adapting asynchrony onto subsequent, physically simultaneous pairs. Thus, in the Miyazaki study, adaptation caused tactile stimuli to feel progressively *less* synchronous. It is possible to speculate that this may be tenable where observers hold no *a priori* assumptions about the physical temporal relationship between sensory inputs (e.g., (Ernst, 2007; Miyazaki, Nozaki, & Nakajima, 2005)), which – given the infrequency of left-right hand index finger stimulation in a natural environment – may be the case for the data of Miyazaki and colleagues. A recent study by Sato et al. acknowledged that variations in the prior distribution could be used to model adaptation effects in multisensory perception (Sato, Toyoizumi, & Aihara, 2007). However, these authors adopted an alternative approach in which adaptation is modelled via changes in the likelihood functions of the senses involved. It remains to be seen which of these two possibilities provides the best quantitative account for the magnitude of adaptation effects across a wide range of multisensory and sensorimotor delays.

Elements of the model described here have parallels with two concepts from the causality literature: firstly, the notion of an ‘internal comparison process’ first postulated by Helmholtz (Helmholtz, 1850) and recently reviewed by Synofzik and colleagues (Synofzik, Vosgerau, & Newen, 2008). In the internal comparison process, observers compute the difference between the predicted and perceived afferent sensory feedback following completion of their motor commands (e.g., a saccade (Lindner, Thier, Kircher, Haarmeier, & Leube,

2005) or contact between fingers (Bays, Flanagan, & Wolpert, 2006)). The output of this ‘comparator model’ gives a metric of causality which is compatible with changes in the perceptual weight allocated to the synchrony prior here: it could be argued that the observers in the present study use the output of an internal comparison process to recalibrate the prior (i.e., change $w_{prior}$) in response to changes in sensorimotor delay. For example, the work of Bays et al. (2006) suggests that the effects reported here may represent a general principal of how prediction is employed in a variety of situations such as tactile force perception. Tactile sensation is typically attenuated when it is associated with self-generated actions (Bays et al., 2005; Shergill, Samson, Bays, Frith, & Wolpert, 2005). Bays and colleagues have shown that this attenuation can occur without the full completion of the action, presumably because the nervous system makes prior assumptions about the sensory consequences of the action (Bays et al., 2006).

Secondly, a synchrony prior also has implications for the nature of intentional binding effects described by Haggard and colleagues, in which observers consistently underestimate the temporal interval between voluntary actions and their sensory consequences (e.g., Haggard et al. (2002)). The prior described in the present study offers a potential explanation for such effects. If, during the adaptation phase, observers had been asked to estimate the size of the delay between action and event (as opposed to their temporal order), it is likely that the interval would be underestimated (Engbert et al., 2008). In Figure 7.11(a), this is shown as the posterior shifting towards zero. Moreover, intentional binding effect size has been shown to reduce with increasing sensorimotor delay (Haggard et al., 2002), a finding predicted by a corresponding reduction in $w_{prior}$. These considerations suggest that intentional binding effects are a manifestation of temporal *realignment* rather than a compression of perceived temporal interval (Haggard et al., 2002; Liddle & Jackson, 2006; Stetson et al., 2006).

Although the precise relationship between intentional binding and the illusion demonstrated here remains speculative, a recent study offers evidence suggesting further similarities with the results reported here. Engbert and colleagues employed an interval estimation task to quantify intentional binding between motor actions and visual, auditory and tactile events; it was observed that intervals terminated by an auditory event were perceived as shorter than those terminated by visual or tactile events, and thus that intentional binding is largest when using auditory events (Engbert et al., 2008). The finding that auditory-motor timing is most flexible is consistent with the data shown in Figure 7.10, which shows that the illusion is greatest in magnitude when the test stimulus is auditory. However, an obvious point of difference between the studies is that intentional binding is measured in the absence of adaptation, unlike the illusion discussed here.

In summary, the novel Bayesian framework presented here suggests that temporal recalibration occurs because actions and their sensory consequences ‘should’ feel synchronous (Kording, 2007). When this *a priori* assumption about the external world is combined with noisy sensorimotor timing estimates, adaptation initiates a realignment of perception away from veridicality and toward the temporal relationship signalled by the prior. Just as observers tend to impose perceptual surface geometries that are commensurate with ‘light must have come from above’ shading patterns (Ramachandran, 1988; Stone, Kerrigan, & Porrill, in press; Sun & Perona, 1998), the imposition of sensorimotor synchrony can bring about realignment of sensory temporal estimates. Importantly, this only occurs when the nervous system can be confident that the sensory consequences are a product of its own motor commands.

Temporal discrepancies between motor actions and sensory events have been shown to be a powerful metric in the perception of causality (Asai & Tanno, 2008; Bays et al., 2005; Franck et al., 2001) and the data presented here imply that the strength of this association declines exponentially with time (Figure 7.7). This makes sense if the nervous system seeks to avoid potentially dangerous recalibration between motor actions and sensory events with independent, external causes. An interesting direction for future work would be to use the paradigm employed in the present work to probe sensorimotor recalibration in schizophrenic patients with delusions of control. The work of two recent studies suggests that the temporal tuning of the effects shown here (Figure 7.7) may be considerably more narrow as a result of their tendency to attribute *external* causalities to *internally* generated stimuli (Lindner et al., 2005; Shergill et al., 2005). By the same token, it would be of interest to examine whether sensorimotor recalibration occurs when observers attempt to interpret actions and sensory consequences, but where the actions are generated by external agencies (Bays et al., 2006; Engbert, Wohlschlager, Thomas, & Haggard, 2007).

# Conclusions

The experiments contained within this thesis contribute to an ever-growing body of psychophysical literature examining the perceived timing of sensory experience. Given the differences in sensory latency between the visual, auditory and tactile systems (as reviewed in Chapter 1), the overwhelmingly veridical percept of multisensory events in a natural environment may be considered surprising. The experiments reported here contain important clues as to the nature of the neural mechanisms mediating perceived timing in humans.

In Chapter 4, estimates of sensory latency derived from reaction time (RT) and temporal order judgment (TOJ) data were compared. The two measures of latency appear unrelated (Section 4.3.4), and this finding cannot be ascribed to differences in the number of modalities attended to (Section 4.4.4) or the lack of any modality identification component to a simple RT task (Section 4.5.4). An implication of these results is that RT and TOJ measure some fundamentally differing aspect(s) of sensory latency. The results are difficult to reconcile with existing models which attempt to explain a variety of dissociations between RT and TOJ, possibly because existing models are primarily developed to account for the differential effects of stimulus intensity on the two measures of latency in a unimodal (rather than crossmodal) context. The lack of relation between RT and TOJ is consistent with the two tasks utilising different information about the stimuli in the different tasks, and/or different underlying neural substrates. It is also shown that dividing attention between sensory modalities has no significant effects on tactile latency (Section 4.4.4), a finding with clear ecological benefits.

The experiments reported in Sections 5.2 and 5.3 investigated the effect of adaptation to asynchronous stimulus pairs on perception of audiovisual (AV), audiotactile (AT) and visuotactile (VT) temporal order. Robust shifts in perceived timing subsequent to adaptation were recorded in all three sensory pairings (Sections 5.2.4 and 5.3.3), contrary to previous hypothesis that temporal recalibration of tactile events does not occur. Close correspondence in the magnitude of temporal recalibration and observer sensitivity to temporal order between the three sensory pairings is suggestive of a single, supramodal neural mechanism mediating the observed effects in all three sensory pairings.

Further analysis of all TOJ data suggests that crossmodal PSS is intransitive, i.e. PSS for the VT modality pairing cannot be inferred from knowledge of PSS in the AV and AT pairings (Section 5.4.3). In isolation, this finding could suggest either that there is no common mechanism for perceived temporal order serving these sensory pairings, or that sensory latency in a crossmodal TOJ task for each modality is dependent upon which modality it is paired with. Crucially, analysis of observer sensitivity revealed no significant differences between the three sensory pairings, in spite of the clear intermodal differences in temporal processing characteristics discussed in Chapter 2. Thus, the more likely conclusion is that a single decisional mechanism mediates crossmodal TOJs, but that the latency of each individual modality is dependent on which modality it is paired with.

In Section 6.4.4, it is shown that perceived AV temporal order can be modulated by adaptation to asynchronous AV stimulus pairs in a spatially-specific manner. This spatial specificity is consistent with an early-stage neural substrate, in apparent contrast to the results reported in Sections 5.2.4 and 5.3.3. However, it may also be consistent with a high-level mechanism which associates the different adapting stimulus polarities with their different locations in space. Should this be the case, it would imply that the observed effects are not truly spatially-specific, but merely reflect the fact that in this experiment the two stimulus pairings were defined by location. Future work should seek to elucidate this crucial detail by investigating whether similarly opposing temporal recalibration can be elicited by defining the stimulus pairs using some factor other than spatial location, for example chromaticity and pitch.

The effect of attention on AV temporal recalibration following adaptation to asynchronous AV stimulus pairs is investigated in Section 6.5. Results show that the magnitude of recalibration is strongly influenced by the focus of observers' attention during the adaptation and 'top-up' phases of the experiment: attending to the temporal order of the adapting stimulus pairs produces significantly more recalibration than attending to the contrast polarity of a fixation cross or the spatial characteristics of the stimulus pair. Thus, attending to the parameter to be judged during the test phase appears to produce the greatest aftereffects, analogous to previous work examining the motion aftereffect (MAE). The results are consistent with a late-stage, post-attentive temporal recalibration mechanism mediating AV asynchrony adaptation.

Chapter 7 presents a series of experiments investigating perceived sensorimotor timing, the results of which indicate that the nervous system recalibrates the perceived temporal relationship between a voluntary motor action and its sensory consequence. A consequence of this is that sensory feedback presented with an unexpectedly short delay appears to precede the motor action which caused it. The data show that this occurs with visual, auditory and tactile events (Section 7.2.4), and that the magnitude of the effect is critically dependent on the adapting delay between action and sensory event (Section 7.3.3). From this, it is concluded that the strength of the causal link between action and event is critical to the illusion. In contrast, the modality of the sensory event appears insignificant – a conclusion strongly supported by the results of Section 7.4.3, which show that changing the modality of the sensory event between adapt and test phases does not modulate the illusion. From this result, it can be concluded that a single supramodal perceptual mechanism mediates the results in the three sensorimotor pairings tested. A novel model is presented which describes the data well. The model proposes a strong prior assumption of physical synchrony between observers' actions and their sensory consequences, the perceptual weighting of which declines exponentially with increasing temporal delay (and correspondingly reduced sense of causality) between action and event.

In conclusion, the experiments described within this thesis further advance our understanding of the factors that influence perceived sensory and sensorimotor timing in humans. It is clear that perceived timing is not dependent merely on transduction and neural latencies (Sections 1.1.1-1.1.3), but is influenced by factors such as the nature of the task employed, level of stimulus certainty, recent experience, and the voluntary nature of our actions. A critical area for future work will be the precise nature of the neural mechanisms underpinning our perception of time.

# List of Acronyms Used

| | |
|---|---|
| AES | Anterior Ectosylvian Sulcus (brain region) |
| AFC | Alternative Forced Choice (e.g., 2AFC) |
| AFF | Auditory Flutter Fusion (threshold) |
| ANOVA | Analysis of Variance |
| AT | Audiotactile |
| AV | Audiovisual |
| BDT | Bayesian Decision Theory |
| CFF | Critical Flicker Fusion (threshold) |
| CRT | Cathode Ray Tube |
| DCML | Dorsal-Column-Medial-Lemniscal (sensory pathway) |
| EEG | Electroencephalography |
| ERP | Event-Related Potential |
| HRTF | Head-Related Transfer Function |
| HSD | Honestly-Significant Difference |
| IID | Interaural Intensity Difference |
| IPD | Interaural Phase Difference |
| ITD | Interaural Time Difference |
| JND | Just-Noticeable Difference |
| LED | Light-Emitting Diode |
| LGN | Lateral Geniculate Nucleus (brain region) |
| LIP | Lateral Intraparietal (brain region) |
| MAE | Motion Aftereffect |

| | |
|---|---|
| MAP | Maximum a Posteriori |
| MEG | Magnetoencephalography |
| MLE | Maximum-Likelihood Estimation |
| MOCS | Method Of Constant Stimuli |
| MRI | Magnetic Resonance Imaging |
| MSE | Modality Shifting Effect |
| NCV | Nerve Conduction Velocity |
| PEST | Parameter Estimation by Sequential Testing |
| PSE | Point of Subjective Equality |
| PSS | Point of Subjective Simultaneity |
| rLS | Lateral Suprasylvian Cortex (brain region) |
| RSE | Redundant Stimulus Effect |
| RT | Reaction Time |
| SAS | Synchronous-Asynchronous Judgment |
| SC | Superior Colliculus (brain region) |
| SD | Standard Deviation |
| SDT | Signal Detection Theory |
| SEM | Standard Error of the Mean |
| SEP | Somatosensory Evoked Potential |
| SJ | Simultaneity Judgment |
| SMA | Supplementary Motor Area (brain region) |
| SOA | Stimulus Onset Asynchrony |
| STS | Superior Temporal Sulcus (brain region) |
| TFF | Two-Flash Flicker (threshold) |
| TMS | Transcranial Magnetic Stimulation |

| | |
|---|---|
| TOJ | Temporal Order Judgment |
| Tpt | Tempero-Parietal Association Cortex (brain region) |
| VEP | Visual Evoked Potential |
| VIP | Ventral Intraparietal (brain region) |
| VSG | Visual Stimulus Generator |
| VT | Visuotactile |

# References

Adelstein, B. D., Begault, D. R., Anderson, M. R., & Wenzel, E. M. (2003). Sensitivity to Haptic-Audio Asynchrony. *Proceedings of the 5th International Conference on Multimodal Interfaces*, 73-76.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257-262.

Alais, D., & Carlile, S. (2005). Synchronising 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*, 2244-2247.

Allan, L. G. (1975). Temporal order psychometric functions based on confidence-rating data. *Perception and Psychophysics, 18*, 369-372.

Allison, T., Goff, W. R., Williamson, P. D., & van Gilder, J. C. (1980). On the neural origin of early components of the human somatosensory evoked potential. In J. E. Desmedt (Ed.), *Progress in Clinical Neurophysiology, Vol. 7: Clinical Uses of Cerebral, Brainstem and Spinal Somatosensory Evoked Potentials* (pp. 51-68). Basel: Karger.

Alvarado, J. C., Stanford, T. R., Vaughan, J. W., & Stein, B. E. (2007). Cortex mediates multisensory but not unisensory integration in superior colliculus. *Journal of Neuroscience, 27*, 12775-12786.

Amano, K., Goda, N., Nishida, S., Ejima, Y., Takeda, T., & Ohtani, Y. (2006). Estimation of the timing of human visual perception from magnetoencephalography. *Journal of Neuroscience, 26*, 3981-3991.

Aminoff, M. J. (2003). Electrophysiology. In C. G. Goetz (Ed.), *Textbook of Clinical Neurology* (2nd ed.). Philadelphia: Saunders.

Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum Likelihood Integration of rapid flashes and beeps. *Neuroscience Letters, 380*, 155-160.

Arezzo, J. C., Vaughan, H. G., & Legatt, A. D. (1981). Topography and intracranial sources of somatosensory evoked potentials in the monkey. II. Cortical components. *Electroencephalography and Clinical Neurophysiology, 51*, 1-18.

Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research, 45*, 1275-1284.

Arrighi, R., Alais, D., & Burr, D. (2005). Neural latencies do not explain the auditory and audio-visual flash-lag effect. *Vision Research, 45*, 2917-2925.

Asai, T., & Tanno, Y. (2007). The relationship between the sense of self-agency and Schizotypal personality traits. *Journal of Motor Behavior, 39*, 162-168.

Asai, T., & Tanno, Y. (2008). Highly schizotypal students have a weaker sense of self-agency. *Psychiatry and Clinical Neurosciences, 62*, 115-119.

Aschersleben, G. (1994). *Afferente Informationen und die Synchronisation von Erieignissen*. Frankfurt, Germany: Lang.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition, 48*, 66-79.

Aschersleben, G., Gehrke, J., & Prinz, W. (2004). A psychophysical approach to action timing. In C. Kaernbach & E. Schroger & H. Muller (Eds.), *Psychophysics beyond sensation: Laws and invariants in human cognition* (pp. 117-136). Hillsdale, NJ: Erlbaum.

Aschersleben, G., Stenneken, P., Cole, J., & Prinz, W. (2002). Timing mechanisms in sensorimotor synchronization, *Common Mechanisms in Perception and Action* (Vol. 19, pp. 227-244).

Avillac, M., Ben Hamed, S., & Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *Journal of Neuroscience, 27*, 1922-1932.

Baker, G. E., & Stryker, M. P. (1990). Retinofugal fibres change conduction velocity and diameter between the optic nerve and tract in ferrets. *Nature, 344*, 342-345.

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of Self-Generated Tactile Sensations Is Predictive, not Postdictive. *PLoS Biology, 4*, e28.

Bays, P. M., Wolpert, D. M., & Flanagan, J. R. (2005). Perception of the Consequences of Self-Action Is Temporally Tuned and Event Driven. *Current Biology, 15*, 1125-1128.

Behar, I., & Bevan, W. (1961). The perceived duration of auditory and visual intervals: cross-modal comparison and interaction. *American Journal of Psychology, 74*, 17-26.

Bernstein, L. E., Auer, E. T., Wagner, M., & Ponton, C. W. (2008). Spatiotemporal dynamics of audiovisual speech processing. *NeuroImage, 39*, 423-435.

Bessel, F. W. (1822). *Astronomische Beobachtungen auf der Koniglichen Universitats-Sternwarte in Konigsberg*. Konigsberg, Germany.

Besser, G. M. (1967). Some Physiological Characteristics of Auditory Flutter Fusion in Man. *Nature 214*, 17-19.

Besser, G. M., Duncan, C., & Quilliam, J. P. (1966). Modification of the Auditory Flutter Fusion Threshold by Centrally Acting Drugs in Man. *Nature 211*, 751.

Blake, R., Martens, W., & Di Gianfilippo, A. (1980). Reaction time as a measure of binocular interaction in human vision. *Investigative Ophthalmology and Visual Science, 19*, 930-941.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience, 11*, 551-559.

Blauert, J. (1997). *Spatial Hearing: the Psychophysics of Human Sound*. London: MIT Press.

Bolz, J., Rosner, G., & Wassle, H. (1982). Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. *Journal of Physiology-London, 328*, 171-190.

Botwinick, J., & Thompson, L. W. (1966). Premotor and Motor Components of Reaction Time. *Journal of Experimental Psychology, 71*, 9-15.

Boulter, L. R. (1977). Attention and reaction times to signals of uncertain modality. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 379-388.

Boutet, I., Rivest, J., & Intriligator, J. (1996). The role of attention on motion, colour and luminance aftereffects. *Investigative Ophthalmology and Visual Science, 37(supplement)*, S528.

Brebner, J. M. T., & Welford, A. T. (1980). Introduction: An Historical Background Sketch. In A. T. Welford (Ed.), *Reaction Times*. London: Academic Press Inc.

Brodal, P. (2004). *The Central Nervous System: Structure and Function* (3rd ed.). New York: Oxford University Press, Inc.

Bruno, A., Ayhan, I., & Johnston, A. (2008). Retinotopic adaptation can influence the apparent duration of a visual stimulus [Abstract]. *Journal of Vision, 8 (6):364a*.

Buchthal, F. (1982). Human nerve potentials evoked by tactile stimuli. I. Maximum conduction and properties of compound potentials. *Acta Physiologica Scandinavica. Supplementum, 502*, 5-18.

Bueti, D., Bahrami, B., & Walsh, V. (2008a). Sensory and association cortex in time perception. *Journal of Cognitive Neuroscience, 20*, 1054-1062.

Bueti, D., van Dongen, E. V., & Walsh, V. (2008b). The Role of Superior Temporal Cortex in Auditory Timing. *PLoS ONE, 3*, e2481.

Bueti, D., Walsh, V., Frith, C., & Rees, G. (2008c). Different brain circuits underlie motor and perceptual representations of temporal intervals. *Journal of Cognitive Neuroscience, 20*, 204-214.

Buonomano, D. V., & Karmarkar, U. R. (2002). How do we tell time? *Neuroscientist, 8*, 42-51.

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience, 10*, 423-425.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21*, 300-304.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11*, 1110-1123.

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*, 2619-2623.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iverson, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*, 191-205.

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience, 22*, 2886-2902.

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*, 1-14.

Caruso, G., Nolano, M., Lullo, F., Crisci, C., Nilsson, J., & Miassini, R. (1994). Median nerve sensory responses evoked by tactile stimulation of the finger proximal and distant phalanx in normal subjects. *Muscle and Nerve, 17*, 269-275.

Cattell, J. M. (1886). The Influence of the Intensity of the Stimulus on the Length of the Reaction Time. *Brain, 8*, 512-515.

Chaston, A., & Kingstone, A. (2004). Time estimation: The effect of cortically mediated attention. *Brain and Cognition, 55*, 286-289.

Chaudhuri, A. (1990). Modulation of the motion aftereffect by selective attention. *Nature, 344*, 60-62.

Choe, C. S., Welch, R. B., Guilford, R. M., & Juola, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception and Psychophysics, 18*, 55-60.

Clark, V. P., & Hillyard, S. A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. *Journal of Cognitive Neuroscience, 8*, 387-402.

Clifford, C. W. G., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision Research, 47*, 3125-3131.

Cobbs, W. H., & Pugh, E. N. (1987). Kinetics and components of the flash photocurrent of isolated retinal rods of the larval salamander, *Ambystoma tigrinum*. *Journal of Physiology-London, 394*, 529-572.

Colavita, F. B. (1974). Human Sensory Dominance. *Perception & Psychophysics, 16*, 409-412.

Corballis, P. M., Funnell, M. G., & Gazzaniga, M. S. (2002). Hemispheric asymmetries for simple visual judgments in the split brain. *Neuropsychologia, 40*, 401-410.

Corey, D. P., & Hudspeth, A. J. (1979). Response latency of vertebrate hair cells. *Biophysical Journal, 26*, 499-506.

Corliss, D. A., & Norton, T. T. (2002). Principles of Psychophysical Measurement. In T. Y. Norton & D. A. Corliss & J. E. Bailey (Eds.), *The Psychophysical Measurement of Visual Function*. Woburn, MA: Butterworth Heinemann.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology, 75*, 485-491.

Correa, A., Sanabria, D., Spence, C., Tudela, P., & Lupianez, J. (2006). Selective temporal attention enhances the temporal resolution of visual perception: Evidence from a temporal order judgment task. *Brain Research, 1070*, 202-205.

Corso, G. M. (1980). Auditory temporal order and perceived fusion-nonfusion. *Perception and Psychophysics, 28*, 465-470.

Corwin, T. R., & Boynton, R. M. (1968). Transitivity of visual judgments of simultaneity. *Journal of Experimental Psychology, 78*, 560-568.

Craig, J. C., & Rollman, G. B. (1999). Somesthesis. *Annual Review of Psychology, 50*, 305-331.

Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science, 12*, 532-535.

Danquah, A. N., Farrell, M. J., & O'Boyle, D. J. (2008). Biases in the subjective timing of perceptual events: Libet et al. (1983) revisited. *Consciousness and Cognition, 17*, 616-627.

Davson, H. (1972). *The physiology of the eye*. Edinburgh & London: Churchill Livingstone.

Deas, R. W., Roach, N. W., & McGraw, P. V. (2008). Distortions of perceived auditory and visual space following adaptation to motion. *Experimental Brain Research, 191*, 473-485.

Di Luca, M., Machulla, T., & Ernst, M. O. (2007). Temporal adaptation influences non-adapted modality pairs. *Perception 36 (ECVP Abstract Supplement)*, 207.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66*, 1388-1404.

Dinnerstein, A. J., & Zlotogura, P. (1968). Intermodal perception of temporal order and motor skills: Effects of age. *Perceptual and Motor Skills, 26*, 987-1000.

Dittrich, W. H., & Henderson, L. (1999). Preparing to react in the absence of uncertainty: II. Stimulus uncertainty and response compatibility in tactile reaction time. *British Journal of Psychology, 90*, 349-372.

Dunlap, K. (1910). Reactions on rhythmic stimuli, with attempt to synchronise. *Psychological Review, 17*, 399-416.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology, 18*, 131-136.

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. *Trends in Cognitive Sciences, 6*, 323-325.

Edwards, S. B., Ginsburgh, C. L., Henkel, C. K., & Stein, B. E. (1979). Sources of subcortical projections to the superior colliculus in the cat. *Journal of Comparative Neurology, 184*, 309-329.

Egeth, H. E., & Sager, L. C. (1977). On the locus of visual dominance. *Perception & Psychophysics, 22*, 77-86.

Ehrensing, R. H., & Lhamon, W. T. (1966). Comparison of tactile and auditory time judgments. *Perceptual and Motor Skills, 23*, 929-930.

Ehret, G., & Scheich, H. (1997). The auditory cortex. *Journal of Comparative Physiology A-Sensory Neural and Behavioral Physiology, 181*, 547-557.

Elliott, R. (1968). Simple visual and simple auditory reaction time: A comparison. *Psychonomic Science, 10*, 335-336.

Engbert, K., & Wohlschlager, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition, 16*, 255-264.

Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition, 107*, 693-704.

Engbert, K., Wohlschlager, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology-Human Perception and Performance, 33*, 1261-1268.

Engel, G. R., & Dougherty, W. G. (1971). Visual-Auditory Distance Constancy. *Nature, 234*, 308.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7(5):7*, 1-14.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162-169.

Exner, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Processe. *Archiv fur die Gesammte Physiologie des Menschen und der Thiere, 11*, 403-432.

Fain, G. L. (2003). *Sensory Transduction*. Sunderland, MA: Sinauer Associates.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of Multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*, 5749-5759.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf & Harterl.

Ferree, C. E., & Rand, G. (1927). Intensity of light and speed of vision studied with special reference to industrial situations. *Transactions of the Illuminating Engineering Society, 22*, 79-110.

Ferry, E. S. (1892). Persistence of vision. *American Journal of Science, XLIV*, 192-207.

Fink, M., Ulbrich, P., Churan, J., & Wittmann, M. (2006). Stimulus-dependent processing of temporal order. *Behavioural Processes, 71*, 344-352.

Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience, 4*, 1171-1173.

Formby, C., Morgan, L. N., Forrest, T. G., & Raney, J. J. (1992). The role of frequency selectivity in measures of auditory and vibrotactile temporal resolution. *The Journal of the Acoustical Society of America, 91*, 293-305.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans - A framework for defining "early" visual processing. *Experimental Brain Research, 142*, 139-150.

Fraisse, P. (1980). Les synchronisations sensori-motrices aux rhythmes. In J. Requin (Ed.), *Anticipation et comportement* (pp. 233-257). Paris, France: Centre National.

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d'Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry, 158*, 454-459.

Frey, R. D. (1990). Selective attention, event perception and the criterion of acceptability principle: evidence supporting and rejecting the doctrine of prior entry. *Human Movement Science, 9*, 481-530.

Frishman, L. J. (2005). Basic Visual Processes. In E. B. Goldstein (Ed.), *Blackwell Handbook of Sensation & Perception*. Malden, MA: Blackwell Publishing.

Froeberg, S. (1907). The relation between magnitude of the stimulus and the time of the reaction. *Archives of Psychology NY, 8*, 25.

Fu, K. M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., Garraghty, P. E., & Schroeder, C. E. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience, 23*, 7510-7515.

Fu, K. M. G., Shah, A. S., O'Connell, M. N., McGinnis, T., Eckholdt, H., Lakatos, P., Smiley, J., & Schroeder, C. E. (2004). Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. *Journal of Neurophysiology, 92*, 3522-3531.

Fucci, D., & Petrosino, L. (1984). Cutaneous temporal resolution for "sophisticated" and "naive" subjects. *Perceptual and Motor Skills, 58*.

Fujisaki, W., & Nishida, S. (2005). Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental Brain Research, 166*, 455-464.

Fujisaki, W., & Nishida, S. (2007a). *Audio-tactile, visuo-tactile, and audio-visual temporal synchrony perception.* Paper presented at the International Multisensory Research Forum, Sydney.

Fujisaki, W., & Nishida, S. (2007b). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research, 47*, 1075-1093.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience, 7*, 773-778.

Funnell, M. G., Corballis, P. M., & Gazzaniga, M. S. (2003). Temporal discrimination in the split brain. *Brain and Cognition, 53*, 218-222.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Psychology, 72*, 521-528.

Geldard, F. A. (1972). *The Human Senses* (2nd ed.). New York: John Wiley & Sons, Inc.

Geldard, F. A., & Sherrick, C. E. (1971). *Cutaneous Temporal Acuity*. Princeton, NJ: Princeton Cutaneous Research Project (report 18).

Gelfand, S. A., Porrazzo, J., & Silman, S. (1988). Aging effects on gap detection thresholds among normal and hearing impaired subjects. *Journal of the Acoustical Society of America, 83*, S75.

Gescheider, G. A. (1966). Resolving of Successive Clicks by the Ears and Skin. *Journal of Experimental Psychology, 71*, 378-381.

Gescheider, G. A. (1967). Auditory and Cutaneous Temporal Resolution of Successive Brief Stimuli. *Journal of Experimental Psychology, 75*, 570-572.

Gescheider, G. A. (1974). Temporal relations in cutaneous stimulation. In F. A. Geldard (Ed.), *Cutaneous communication systems and devices*. Austin, TX: The Psychonomic Society.

Gibbon, J., & Rutschmann, R. (1969). Temporal Order Judgment and Reaction Time. *Science, 47*, 413-415.

Gielen, S., Schmidt, R. A., & Vandenheuvel, P. J. M. (1983). On the nature of intersensory facilitation of reaction time. *Perception and Psychophysics, 34*, 161-168.

Gobbelé, R., Schürmann, M., Forss, N., Juottonen, K., Buchner, H., & Hari, R. (2003). Activation of the human posterior parietal and temporoparietal cortices during audiotactile interaction. *NeuroImage, 20*, 503-511.

Goldring, S., Aras, E., & Weber, P. C. (1970). Comparative study of sensory input to motor cortex in animals and man. *Electroencephalography and Clinical Neurophysiology, 29*, 537-550.

Goldstein, E. B. (1999). *Sensation and Perception* (5th ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Goldstone, S. (1968). Reaction time to onset and termination of lights and sounds. *Perceptual and Motor Skills, 27*, 1023-1029.

Goldstone, S., & Goldfarb, J. L. (1963). Judgment of filled and unfilled durations: Intersensory factors. *Perceptual and Motor Skills, 17*, 763-774.

Gomes, G. (1998). The timing of conscious experience: A critical review and reinterpretation of Libet's research. *Consciousness and Cognition, 7*, 559-595.

Goodfellow, L. D. (1933). An empirical comparison of audition, vision, and touch in the discrimination of short intervals of time. *American Journal of Psychology, 45*, 243-258.

Granit, R., & Harper, P. (1930). Comparative studies on the peripheral and central retina. *American Journal of Physiology, 95*, 211-228.

Grantham, D. W. (1995). Spatial Hearing and Related Phenomena. In B. C. J. Moore (Ed.), *Hearing*. London: Academic Press.

Gray, J. A. B., & Sato, M. (1953). Properties of the receptor potential in Pacinian corpuscles. *Journal of Physiology 122*, 610-636.

Green, D. M. (1971). Temporal auditory acuity. *Psychological Review, 78*, 540-551.

Gregory, R. L. (1967). Origin of Eyes and Brains. *Nature 213*, 369-372.

Grondin, S., & Girard, C. (2005). About hemispheric differences in the processing of temporal intervals. *Brain and Cognition, 58*, 125-132.

Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception and Psychophysics, 49*, 245-256.

Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hahnel, S., Picton, T. W., & Scherg, M. (1999). Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. *Clinical Neurophysiology, 110*, 856-868.

Hackett, T. A., & Kaas, J. H. (2003). Auditory Processing in the Primate Brain. In M. Gallagher & R. J. Nelson (Eds.), *Handbook of Psychology Vol. 3: Biological Psychology*. New York: John Wiley & Sons, Inc.

Hackney, C. M. (1987). Anatomical features of the auditory pathway from cochlea to cortex. *British Medical Bulletin, 43*, 780-801.

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition, 12*, 695-707.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*, 382-385.

Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. (2003). Awareness of action in schizophrenia. *NeuroReport, 14*, 1081-1085.

Hallett, M. (2007). Volitional control of movement: The physiology of free will. *Clinical Neurophysiology, 118*, 1179-1192.

Hamlin, A. J. (1895). On the least observable interval between stimuli addressed to disparate senses and to different organs of the same sense. *American Journal of Psychology, 6*, 564-575.

Handy, T. C., Gazzaniga, M. S., & Ivry, R. B. (2003). Cortical and subcortical contributions to the representation of temporal information. *Neuropsychologia, 41*, 1461-1473.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*, 1367-1368.

Harrar, V., & Harris, L. R. (2005). Simultaneity constancy: detecting events with touch and vision. *Experimental Brain Research, 166*, 465-473.

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research, 186*, 517-524.

Harrington, D. L., Haaland, K. Y., & Knight, R. T. (1998). Cortical networks underlying mechanisms of time perception. *Journal of Neuroscience, 18*, 1085-1095.

Hartcher-O'Brien, J., Gallace, A., Krings, B., Koppen, C., & Spence, C. (2008). When vision 'extinguishes' touch in neurologically-normal people: extending the Colavita visual dominance effect. *Experimental Brain Research, 186*, 643-658.

Hartmann, E., Lachenmayr, B., & Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research, 19*, 1019-1023.

Harvey, L. O., Jr. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments & Computers, 18*, 623-632.

Hawkes, G. R., Deardorff, P. A., & Ray, W. S. (1977). Response delay effects with cross-modality duration judgments. *Journal of Auditory Research, 17*, 55-57.

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research, 193*, 307-314.

Hecht, D., Reiner, M., & Karni, A. (2008). Multisensory enhancement: gains in choice and in simple response times. *Experimental Brain Research, 189*, 133-143.

Hecht, S., & Verrip, C. D. (1933). Intermittent stimulation by light III: The relation between intensity and critical flicker fusion frequency for different retinal regions. *Journal of General Physiology, 17*, 251.

Helmholtz, H. L. F. (1850). Messingen über den zeitlichen Verlanf der Zucking animalischer. Muskeln und die Fortpflanzungsgeschwindigkeit der Reizung in den Nerven. *Archiv für Anatomie und Physiologie*, 276-364.

Helson, H. (1964). *Adaptation-Level Theory*. New York: Harper & Row.

Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research, 44*, 2875-2884.

Heron, J., Whitaker, D., McGraw, P. V., & Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision, 7(13):5*, 1-8.

Herrick, R. M. (1974). Frequency thresholds for two-flash flicker and critical flicker: Why they differ. *Perception and Psychophysics, 15*, 79-82.

Hershenson, M. (1962). Reaction Time as a Measure of Intersensory Facilitation. *Journal of Experimental Psychology, 63*, 289-293.

Hestrin, S., & Korenbrot, J. I. (1990). Activation kinetics of retinal cones and rods: Response to intense flashes of light. *Journal of Neuroscience, 10*, 1967-1973.

Hicks, T. P., Benedek, G., & Thurlow, G. A. (1988). Modality Specificity of Neuronal Responses Within the Cat's Insula. *Journal of Neurophysiology, 60*, 422-437.

Hildreth, J. D. (1973). Bloch's law and a temporal integration model for simple reaction time to light. *Perception and Psychophysics, 14*, 421-432.

Hirsh, I. J. (1959). Auditory Perception of Temporal Order. *Journal of the Acoustical Society of America, 31*, 759-767.

Hirsh, I. J., & Sherrick, C. E. (1961). Perceived Order in Different Sense Modalities. *Journal of Experimental Psychology, 62*, 423-432.

Ho, C., Spence, C., & Tan, H. Z. (2005). Warning Signals Go Multisensory. In N. J. Mahwah (Ed.), *Proceedings of the 11th International Conference on Human-Computer Interaction* (Vol. 9). Las Vegas, Nevada, U.S.A.: Lawrence Erlbaum Associates.

Hodinott-Hill, I., Thilo, K. V., Cowey, A., & Walsh, V. (2002). Auditory Chronostasis: Hanging on the Telephone. *Current Biology, 12*, 1779-1781.

Hohnsbein, J., Falkenstein, M., Hoormann, J., & Blanke, L. (1991). Effects of Crossmodal Divided Attention on Late Erp Components .1. Simple and Choice Reaction Tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 438-446.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia, 45*, 3340-3345.

Holmes, N. P. (2008). *The Seemingly Inviolable Principle of Inverse Effectiveness: In Search of a Null Hypothesis.* Paper presented at the International Multisensory Research Forum, Hamburg.

Hoshiyama, M., Kakigi, R., & Tamura, Y. (2004). Temporal discrimination threshold on various parts of the body. *Muscle and Nerve, 29*, 243-247.

Hseih, S. (2002). Task Shifting in Dual-Task Settings. *Perceptual and Motor Skills, 94*, 407-414.

Hsiao, S. S., Johnson, K., & Yoshioka, T. (2003). Processing of Tactile Information in the Primate Brain. In M. Gallagher & R. J. Nelson (Eds.), *Handbook of Psychology vol. 3: Biological Psychology*. New York: John Wiley & Sons, Inc.

Huerta, M. F., & Harting, J. K. (1984). The mammalian superior colliculus: Studies of its morphology and connections. In H. Vanegas (Ed.), *Comparative neurology of the optic tectum*. New York: Plenum Press.

Hume, D. (1748). *Philosophical Essays Concerning Human Understanding*. London: A. Millar.

Inui, K., Okamoto, H., Miki, K., Gunji, A., & Kakigi, R. (2006). Serial and parallel processing in the human auditory cortex: A magnetoencephalographic study. *Cerebral Cortex, 16*, 18-30.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and Production of Temporal Intervals across a Range of Durations: Evidence for a Common Timing Mechanism. *Journal of Experimental Psychology-Human Perception and Performance, 21*, 3-18.

Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of Cognitive Neuroscience, 1*, 136-152.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences, 12*, 273-280.

Jahanshahi, M., Jones, C. R. G., Dirnberger, G., & Frith, C. D. (2006). The substantia nigra pars compacta and temporal processing. *Journal of Neuroscience, 26*, 12266-12273.

Jaskowski, P. (1992). Temporal-Order Judgment and Reaction-Time for Short and Long Stimuli. *Psychological Research-Psychologische Forschung, 54*, 141-145.

Jaskowski, P. (1993). Temporal-Order Judgment and Reaction-Time to Stimuli of Different Rise Times. *Perception, 22*, 963-970.

Jaskowski, P. (1996). Simple reaction time and perception of temporal order: Dissociations and hypotheses. *Perceptual and Motor Skills, 82*, 707-730.

Jaskowski, P., Jaroszyk, F., & Hojan-Jezierska, D. (1990). Temporal-Order Judgments and Reaction-Time for Stimuli of Different Modalities. *Psychological Research-Psychologische Forschung, 52*, 35-38.

Jaskowski, P., & Sobieralska, K. (2004). Effect of stimulus intensity on manual and saccadic reaction time. *Perception and Psychophysics, 66*, 535-544.

Jeeves, M. A., & Moes, P. (1996). Interhemispheric transfer time differences related to aging and gender. *Neuropsychologia, 34*, 627-636.

Jiang, W., Jiang, H., & Stein, B. E. (2002). Two corticotectal areas facilitate multisensory orientation behavior. *Journal of Cognitive Neuroscience, 14*, 1240-1255.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology, 85*, 506-522.

Johnson, A. M., Reed, T. E., & Vernon, P. A. (2005). Brain nerve conduction velocity is a valid and useful construct for studying human cognitive abilities: A reply to Saint-Amour et al. *Neuropsychologia, 43*, 1845-1846.

Johnson, K. O. (2001). The roles and functions of cutaneous mechanoreceptors. *Current Opinion in Neurobiology, 11*, 455-461.

Johnson, K. O., Yoshioka, T., & Vega-Bermudez, F. (2000). Tactile functions of mechanoreceptive afferents innervating the hand. *Journal of Clinical Neurophysiology, 17*, 539-558.

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology, 16*, 472-479.

Jueptner, M., Rijntjes, M., Weiller, C., Faiss, J. H., Timmann, D., Mueller, S. P., & Diener, H. C. (1995). Localisation of a cerebellar timing process using PET. *Neurology, 45*, 1540-1545.

Kagerer, F. A., Wittmann, M., Szelag, E., & Steinbüchel, N. v. (2002). Cortical involvement in temporal reproduction: evidence for differential roles of the hemispheres. *Neuropsychologia, 40*, 357-366.

Kammer, T., Lehr, L., & Kirschfeld, K. (1999). Cortical visual processing is temporally dispersed by luminance in human subjects. *Neuroscience Letters, 263*, 133-136.

Kanabus, M., Szelag, E., Rojek, E., & Poppel, E. (2002). Temporal order judgement for auditory and visual stimuli. *Acta Neurobiologiae Experimentalis, 62*, 263-270.

Kanai, R., Paffen, C. L. E., Hogendoorn, H., & Verstraten, F. A. J. (2006). Time dilation in dynamic visual display. *Journal of Vision, 6*, 1421-1430.

Kanai, R., & Watanabe, M. (2006). Visual onset expands subjective time. *Perception & Psychophysics, 68*, 1113-1123.

Kansaku, K., Hanakawa, T., Wu, T., & Hallett, M. (2004). A shared neural network for simple reaction time. *NeuroImage, 22*, 904-911.

Kaplan, H. L. (1975). The five distractors experiment: Exploring the critical band with contaminated white noise. *Journal of the Acoustical Society of America, 58*, 404-411.

Karhu, J., & Tesche, C. D. (1999). Simultaneous Early Processing of Sensory Input in Human Primary (SI) and Secondary (SII) Somatosensory Cortices. *Journal of Neurophysiology, 81*, 2017-2025.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning and Memory, 10*, 141-147.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: Encoding time in neural network states. *Neuron, 53*, 427-438.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research, 167*, 635-640.

Keetels, M., & Vroomen, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research 182*, 559-565.

Keetels, M., & Vroomen, J. (2008). Temporal recalibration to tactile-visual asynchronous stimuli. *Neuroscience Letters, 430*, 130-134.

Keller, P. E., Wascher, E., Prinz, W., Waszak, F., Koch, I., & Rosenbaum, D. A. (2006). Differences between intention-based and stimulus-based actions. *Journal of Psychophysiology, 20*, 9-20.

Kelly, D. H. (1961). Visual response to time-dependent stimuli I: Amplitude sensitivity measurements. *Journal of the Optical Society of America 51*, 422-429.

Kelly, D. L., Goldring, S., & O'Leary, J. L. (1965). Averaged evoked somatosensory responses from exposed cortex of man. *Archives of Neurology, 13*, 1-9.

Kietzman, M. L. (1967). Two-Pulse Measures of Temporal Resolution as a Function of Stimulus Energy. *Journal of the Optical Society of America 57*, 809-813.

King, A. J., & Palmer, A. R. (1985). Integration of visual and auditory information in bimodal neurons in the guinea-pig superior colliculus. *Experimental Brain Research, 60*, 492-500.

King, H. E. (1962). Two-flash and flicker fusion thresholds for normal and schizophrenic subjects. *Perceptual and Motor Skills, 14*, 517-518.

Klein, R. M. (1977). Attention and Visual Dominance - A Chronometric Analysis. *Journal of Experimental Psychology-Human Perception and Performance, 3*, 365-378.

Klein, S. (2002). Libet's Research on the Timing of Conscious Intention to Act: A Commentary. *Consciousness and Cognition, 11*, 273-279.

Klingaman, R. L., & Anch, A. M. (1972). Human auditory evoked response intensity functions from monopolar and bipolar scalp recordings *Psychonomic Science, 29*, 17-19.

Knox, G. W. (1945). Investigations of flicker and fusion: IV. The effect of auditory flicker on the pronouncedness of visual flicker. *Journal of General Psychology, 33*, 145-154.

Kohfeld, D. L. (1971). Simple Reaction Time as a Function of Stimulus Intensity in Decibels of Light and Sound. *Journal of Experimental Psychology, 88*, 251-257.

Kohfeld, D. L., Santee, J. L., & Wallace, N. D. (1981). Loudness and reaction time: I. *Perception and Psychophysics, 29*, 535-549.

Kolers, P. A., & Brewster, J. M. (1985). Rhythms and Responses. *Journal of Experimental Psychology-Human Perception and Performance, 11*, 150-167.

Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception, 33*, 1049-1060.

Koppen, C., & Spence, C. (2007a). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters, 418*, 266-271.

Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research, 1186*, 224-232.

Koppen, C., & Spence, C. (2007c). Seeing the light: exploring the Colavita visual dominance effect. *Experimental Brain Research, 180*, 737-754.

Koppen, C., & Spence, C. (2007d). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters, 417*, 107-111.

Kording, K. (2007). Decision theory: What "should" the nervous system do? *Science, 318*, 606-610.

Kruger, L. (1996). *Pain and Touch*. San Diego: Academic Press.

Lacruz, F., Artieda, J., Pastor, M. A., & Obeso, J. A. (1991). The anatomical basis of somesthetic temporal discrimination in humans. *Journal of Neurology Neurosurgery and Psychiatry, 54*, 1077-1081.

Lakatos, P., Pincze, Z., Fu, K. M. G., Javitt, D. C., Karmos, G., & Schroeder, C. E. (2005). Timing of pure tone and noise-evoked responses in macaque auditory cortex. *NeuroReport, 16*, 933-937.

Laming, D. (1997). *The Measurement of Sensation*. Oxford: Oxford University Press.

Landis, C. (1954). Determinants of the critical flicker fusion threshold. *Physiological Reviews, 34*, 259-286.

Lankheet, M. J. M., & Verstraten, F. A. J. (1995). Attentional modulation of adaptation to two-component transparent motion. *Vision Research, 35*, 1401-1412.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to intention. *Science, 303*, 1208-1210.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. *Journal of Neuroscience, 26*, 7265-7271.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience, 14*, 420-429.

Leshowitz, B. (1971). Measurement of the Two-Click Threshold. *Journal of the Acoustical Society of America, 49*, 462-466.

Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics, 2*, 164-168.

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. *Neuroscience Letters, 357*, 119-122.

Lewis, M. F. (1967). Two-Flash Thresholds as a Function of Luminance in the Dark-Adapted Eye. *Journal of the Optical Society of America, 57*, 814-815.

Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology, 13*, 250-255.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of Conscious Intention to Act in Relation to Onset of Cerebral-Activity (Readiness-Potential) - the Unconscious Initiation of a Freely Voluntary Act. *Brain, 106*, 623-642.

Liddle, E., & Jackson, S. R. (2006). Recalibrating time: When did I do that? *Current Biology, 16*, R994-R996.

Lindner, A., Thier, P., Kircher, T. T. J., Haarmeier, T., & Leube, D. T. (2005). Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology, 15*, 1119-1124.

Lopez-Moliner, J., & Linares, D. (2006). The flash-lag effect is reduced when the flash is perceived as a sensory consequence of our action. *Vision Research, 46*, 2122-2129.

Lutkenhoner, B., Lammertmann, C., Simoes, C., & Hari, R. (2002). Magnetoencephalographic correlates of audiotactile interaction. *NeuroImage, 15*, 509-522.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*, 1206-1208.

Macaluso, E., Frith, C. D., & Driver, J. (2002). Crossmodal spatial influences of touch on extrastriate visual areas take current gaze direction into account. *Neuron, 34*, 647-658.

Macefield, G., Gandevia, S. C., & Burke, D. (1989). Conduction velocities of muscle and cutaneous afferents in the upper and lower limbs in human subjects. *Brain, 112*, 1519-1532.

Machulla, T., Di Luca, M., & Ernst, M. O. (2007). Perception of crossmodal simultaneity is not transitive. *Perception 36*, S209.

Mamassian, P., Landy, M., & Maloney, L. T. (2002). Bayesian Modelling of Visual Perception. In R. P. N. Rao & B. A. Olshausen & M. S. Lewicki (Eds.), *Probabilistic Models of the Brain: Perception and Neural Function* (pp. 13-36). Cambridge, MA: MIT Press.

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics, 11*, 431-441.

Marzi, C. A. (2004). Two brains, one clock. *Trends in Cognitive Sciences, 8*, 1-3.

Mather, G., Verstraten, F., & Anstis, S. (Eds.). (1998). *The Motion Aftereffect*. Cambridge, MA: MIT Press.

Matsuhashi, M., Ikeda, A., Ohara, S., Matsumoto, R., Yamamoto, J., Takayama, M., Satow, T., Begum, T., Usui, K., Nagamine, T., Mikuni, N., Takahashi, J., Miyamoto, S., Fukuyama, H., & Shibasaki, H. (2004). Multisensory convergence at human temporo-parietal junction - epicortical recording of evoked responses. *Clinical Neurophysiology, 115*, 1145-1160.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience, 27*, 307-340.

Maunsell, J. H. R., & Gibson, J. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68*, 1332-1343.

McDonald, J. J., Teder-Salejarvi, W. A., Di Russo, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience, 8*, 1197-1202.

McFarland, R. A., Warren, A. B., & Karis, C. (1958). Alterations in critical flicker frequency as a function of age and light:dark ratio. *Journal of Experimental Psychology, 56*, 529-538.

McKeefry, D. J., Parry, N. R. A., & Murray, I. J. (2003). Simple reaction times in color space: The influence of chromaticity, contrast, and cone opponency. *Investigative Ophthalmology & Visual Science, 44*, 2267-2276.

Meegan, D. V., Aslin, R. N., & Jacobs, R. A. (2000). Motor timing learned without motor training. *Nature Neuroscience, 3*, 860-862.

Meese, T. S. (1995). Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Perception and Psychophysics, 57*, 267-281.

Meijers, L. M. M., & Eijkman, E. G. J. (1977). Distributions of simple RT with single and double stimuli. *Perception and Psychophysics, 22*, 41-48.

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research, 14*, 31-40.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of Multisensory Integration in Superior Colliculus Neurons .1. Temporal Factors. *Journal of Neuroscience, 7*, 3215-3229.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*, 389-391.

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*, 640-662.

Meyer-Lindenberg, A., Miletich, R. S., Kohn, P. D., Esposito, G., Carson, R. E., Quarantelli, M., Weinberger, D. R., & Berman, K. F. (2002). Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. *Nature Neuroscience, 5*, 267-271.

Middlebrooks, J. C., & Green, D. M. (1991). Sound Localization by Human Listeners. *Annual Review of Psychology, 42*, 135-159.

Miller, G. A., & Taylor, W. G. (1948). The Perception of Repeated Bursts of Noise. *The Journal of the Acoustical Society of America, 20*, 171-182.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14*, 247-279.

Miller, J., & Schwarz, W. (2006). Dissociations between reaction times and temporal order judgments: A diffusion model approach. *Journal of Experimental Psychology-Human Perception and Performance, 32*, 394-412.

Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology, 46*, 101-151.

Mitrani, L., Shekerdjiiski, S., & Yakimoff, N. (1986). Mechanisms and Asymmetries in Visual Perception of Simultaneity and Temporal Order. *Biological Cybernetics, 54*, 159-165.

Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian models of human coincidence timing. *Journal of Neurophysiology, 94*, 395-399.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience, 9*, 875-877.

Mizobuchi, K., Kuwabara, S., Toma, S., Nakajima, Y., Ogawara, K., & Hattori, T. (2000). Single unit responses of human cutaneous mechanoreceptors to air-puff stimulation. *Clinical Neurophysiology, 111*, 1577-1581.

Mizobuchi, K., Kuwabara, S., Toma, S., Nakajima, Y., Ogawara, K., & Hattori, T. (2002). Properties of human skin mechanoreceptors in peripheral neuropathy. *Clinical Neurophysiology, 113*, 310-315.

Mollon, J. D., & Perkins, A. J. (1996). Errors of judgment at Greenwich in 1796. *Nature 380*, 101-102.

Moore, B. C. J. (2005). Basic Auditory Processes. In E. B. Goldstein (Ed.), *Blackwell Handbook of Sensation and Perception*. Marden, MA: Blackwell Publishing Ltd.

Moore, D. R. (1987). Physiology of higher auditory system. *British Medical Bulletin, 43*, 856-870.

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience, 8*, 950-954.

Murase, M., Saito, D. N., Kochiyama, T., Tanabe, H. C., Tanaka, S., Harada, T., Aramaki, Y., Honda, M., & Sadato, N. (2008). Cross-modal integration during vowel identification in audiovisual speech: A functional magnetic resonance imaging study. *Neuroscience Letters, 434*, 71-76.

Myers, A. K., Cotton, B., & Hilp, H. A. (1981). Matching the rate of concurrent tone bursts and light flashes as a function of flash surround luminance. *Perception and Psychophysics, 30*, 33-38.

N'Diaye, K., Ragot, R., Garnero, L., & Pouthas, V. (2004). What is common to brain activity evoked by the perception of visual and auditory filled durations? A study with MEG and EEG co-recordings. *Cognitive Brain Research, 21*, 250-268.

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *Journal of Neuroscience, 18*, 1559-1570.

Nagy, A., Eordegh, G., Paroczy, Z., Markus, Z., & Benedek, G. (2006). Multisensory integration in the basal ganglia. *European Journal of Neuroscience, 24*, 917-924.

Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neuroscience Letters, 413*, 72-76.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research, 25*, 499-507.

Neumann, O., Esselmann, U., & Klotz, W. (1993). Differential effects of visual-spatial attention on response latency and temporal-order judgment. *Psychological Research-Psychologische Forschung, 56*, 26-34.

Neumann, O., Koch, R., Niepel, M., & Tappe, T. (1992). Reaction-Time and Temporal-Order Judgments - Agreement or Dissociation. *Zeitschrift Fur Experimentelle Und Angewandte Psychologie, 39*, 621-645.

Nijhawan, R. (2002). Neural delays, visual motion and the flash-lag effect. *Trends in Cognitive Sciences, 6*, 387-393.

Nissen, M. J. (1977). Stimulus intensity and information processing. *Perception and Psychophysics, 22*, 338-352.

Nissen, M. J., & Pokorny, J. (1977). Wavelength effects on simple reaction time. *Perception and Psychophysics, 22*, 457-462.

Nittono, H., Shimizu, K., & Hori, T. (2004). Subjectively optimal delay in computer response to a user's voluntary action. *Perceptual and Motor Skills, 99*, 924-930.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience, 27*, 11431-11441.

Noguchi, Y., & Kakigi, R. (2006). Time representations can be made from nontemporal information in the brain: An MEG study. *Cerebral Cortex, 16*, 1797-1808.

Occelli, V., Hartcher-O'Brien, J., Spence, C., & Zampini, M. (2008a). *Is the Colavita effect an exclusively visual phenomenon?* Paper presented at the International Multisensory Research Forum, Hamburg.

Occelli, V., Spence, C., & Zampini, M. (2008b). Audiotactile temporal order judgments in sighted and blind individuals. *Neuropsychologia, 46*, 2845-2850.

Oyster, C. W. (1999). *The Human Eye*. Sunderland, MA: Sinauer Associates, Inc.

Paillard, J. (1949). Quelques donnees psychophysiologiques relatives au declenchement de la command motrice. *L'Annee Psychologique, 48*, 28-47.

Palmer, A. R. (1995). Neural Signal Processing. In B. C. J. Moore (Ed.), *Hearing*. London: Academic Press.

Pariyadath, V., & Eagleman, D. (2007). The Effect of Predictability on Subjective Duration. *PLoS ONE, 2*, e1264.

Pariyadath, V., & Eagleman, D. (2008). Brief subjective durations contract with repetition. *Journal of Vision, 8(16):11*, 1-6.

Park, J., Schlag-Rey, M., & Schlag, J. (2003). Voluntary action expands perceived duration of its sensory consequence. *Experimental Brain Research, 149*, 527-529.

Pasternak, T., Bisley, J. W., & Calkins, D. (2003). Visual Information Processing in the Primate Brain. In M. Gallagher & R. J. Nelson (Eds.), *Handbook of Psychology vol. 3: Biological Psychology*. New York: John Wiley & Sons, Inc.

Pastor, M. A., Day, B. L., Macaluso, E., Friston, K. J., & Frackowiak, R. S. J. (2004). The functional neuroanatomy of temporal discrimination. *Journal of Neuroscience, 24*, 2585-2591.

Patestas, M. A., & Gartner, L. P. (2006). *A Textbook of Neuroanatomy*. Malden, MA: Blackwell.

Penner, M. J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *Journal of the Acoustical Society of America, 61*, 552-557.

Pentland, A. (1980). Maximum-Likelihood Estimation - The Best PEST. *Perception and Psychophysics, 28*, 377-379.

Perrott, D. R., & Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *The Journal of the Acoustical Society of America, 87*, 1728-1731.

Pesavento, M. J., & Schlag, J. (2006). Transfer of learned perception of sensorimotor simultaneity. *Experimental Brain Research, 174*, 435-442.

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 Section, 96*, 248-260.

Petrosino, L., & Fucci, D. (1985). Stability of temporal resolution of the tactile sensory system for a one-year interim. *Perceptual and Motor Skills, 60*, 590.

Petrosino, L., & Fucci, D. (1989). Temporal resolution of the aging tactile sensory system. *Perceptual and Motor Skills, 68*, 288-290.

Pike, G., & Edgar, G. (2005). Perception. In N. Braisby & A. Gellatly (Eds.), *Cognitive Psychology*. Oxford: Oxford University Press.

Pins, D., & Bonnet, C. (2000). The Pieron function in the threshold region. *Perception & Psychophysics, 62*, 127-136.

Plomp, R. (1964). Rate of Decay of Auditory Sensation. *The Journal of the Acoustical Society of America, 36*, 277-282.

Pockett, S., & Miller, A. (2007). The rotating spot method of timing subjective events. *Consciousness and Cognition, 16*, 241-254.

Poliakoff, E., Shore, D. I., Lowe, C., & Spence, C. (2006). Visuotactile temporal order judgments in ageing. *Neuroscience Letters, 396*, 207-211.

Poppel, E., Schill, K., & von Steinbuchel, N. (1990). Sensory integration within temporally neutral systems states: a hypothesis. *Naturwissenschaften, 77*, 89-91.

Porter, T. C. (1902). Contributions to the study of flicker. *Proceedings of the Royal Society of London, LXX*, 313-329.

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual Dominance - Information-Processing Account of Its Origins and Significance. *Psychological Review, 83*, 157-171.

Ptito, M., Fumal, A., de Noordhout, A. M., Schoenen, J., Gjedde, A., & Kupers, R. (2008). TMS of the occipital cortex induces tactile sensations in the fingers of blind Braille readers. *Experimental Brain Research, 184*, 193-200.

Raab, D. H. (1962a). Effects of stimulus-duration on auditory reaction-time. *American Journal of Psychology, 75*, 298-301.

Raab, D. H. (1962b). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24*, 574-590.

Rains, J. D. (1963). Signal luminance and position effects in human reaction time. *Vision Research, 3*, 239-251.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331*, 163-166.

Rao, A., Nobre, A. C., Alexander, I., & Cowey, A. (2007). Auditory evoked visual awareness following sudden ocular blindness: an EEG and TMS investigation. *Experimental Brain Research, 176*, 288-298.

Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience, 4*, 317-323.

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine, 13*, 214-232.

Recanzone, G. H. (2003). Auditory Influences on Visual Temporal Rate Perception. *Journal of Neurophysiology, 89*, 1078-1093.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16*, 259-272.

Reed, T. E., Vernon, P. A., & Johnson, A. M. (2004). Sex difference in brain nerve conduction velocity in normal humans. *Neuropsychologia, 42*, 1709-1714.

Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.

Repp, B. H. (1999). Control of expressive and metronomic timing in pianists. *Journal of Motor Behavior, 31*, 145-164.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12*, 969-992.

Repp, B. H. (2006). Rate Limits of Sensorimotor Synchronization. *Advances in Cognitive Psychology, 2*, 163-181.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology-Human Perception and Performance, 28*, 1085-1099.

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research-Psychologische Forschung, 68*, 252-270.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B-Biological Sciences, 273*, 2159-2168.

Robinson, E. S. (1934). Work of the Integrated Organism. In C. Murchison (Ed.), *Handbook of General Experimental Psychology*. Worcester, Massachusetts: Clark University Press.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50*, 19-26.

Rodieck, R. W. (1998). *The First Steps in Seeing*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Rolke, B., Ulrich, R., & Bausenhart, K. M. (2006). Attention delays perceived stimulus offset. *Vision Research, 46*, 2926-2933.

Romei, V., Murray, M. M., Merabet, L. B., & Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *Journal of Neuroscience, 27*, 11465-11472.

Rose, D. (2006). Psychophysical Methods. In G. M. Breakwell & S. Hammond & C. Fife-Schaw & J. A. Smith (Eds.), *Research Methods in Psychology* (Third edition ed., pp. 160-181). London: Sage Publications.

Rose, D., Bradshaw, M. F., & Hibbard, P. B. (2003). Attention affects the stereoscopic depth aftereffect. *Perception, 32*, 635-640.

Rose, D., & Summers, J. (1995). Duration illusions in a train of visual stimuli. *Perception, 24*, 1177-1187.

Rose, R. M., Teller, D. Y., & Rendlema, P. (1970). Statistical properties of staircase estimates. *Perception and Psychophysics, 8*, 199-204.

Rosner, B. S. (1961). Neural factors limiting cutaneous spatiotemporal discrimination. In W. R. Rosenblith (Ed.), *Sensory Communication*. Cambridge, MA: MIT Press.

Roufs, J. A. J. (1974). Dynamic Properties of Vision .5. Perception Lag and Reaction-Time in Relation to Flicker and Flash Thresholds. *Vision Research, 14*, 853-869.

Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception and Psychophysics, 34*, 541-548.

Rovamo, J., & Raninen, A. (1984). Critical flicker frequency and M-scaling of stimulus size and retinal illuminance. *Vision Research, 24*, 1127-1131.

Rovamo, J., & Raninen, A. (1988). Critical flicker frequency as a function of stimulus area and luminance at various eccentricities in human cone vision: A revision of Granit-Harper and Ferry-Porter laws. *Vision Research, 28*, 785-790.

Rowland, B. A., Quessy, S., Stanford, T. R., & Stein, B. E. (2007). Multisensory integration shortens physiological response latencies. *Journal of Neuroscience, 27*, 5879-5884.

Rutschmann, J., & Link, R. (1964). Perception of Temporal Order of Stimuli Differing in Sense Mode and Simple Reaction Time. *Perceptual and Motor Skills, 18*, 345-352.

Saint-Amour, D., Saron, C. D., Schroeder, C. E., & Foxe, J. J. (2005). Can whole brain nerve conduction velocity be derived from surface-recorded visual evoked potentials?: A re-examination of Reed, Vernon, and Johnson (2004). *Neuropsychologia, 43*, 1838-1844.

Salami, M., Itami, C., Tsumoto, T., & Kimura, F. (2003). Change of conduction velocity by regional myelination yields constant latency irrespective of distance between thalamus and cortex. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 6174-6179.

Sanford, A. J. (1974). Attention Bias and Relation of Perception Lag to Simple Reaction-Time. *Journal of Experimental Psychology, 102*, 443-446.

Sanford, E. C. (1888a). Personal equation. *American Journal of Psychology, 2*, 3-38.

Sanford, E. C. (1888b). Personal equation III: The nature and cause of personal equation. *American Journal of Psychology, 2*, 403-430.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation, 19*, 3335-3355.

Schmolesky, M. T., Wang, Y. C., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79*, 3272-3278.

Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology, 47*, 333-366.

Schroeder, C. E., Molholm, S., Lakatos, P., Ritter, W., & Foxe, J. J. (2004). Human-Simian correspondence in the early cortical processing of multisensory cues. *Cognitive Processing, 5*, 140-151.

Schroeder, C. E., Seto, S., Arezzo, J. C., & Garraghty, P. E. (1995). Electrophysiological evidence for overlapping dominant and latent inputs to somatosensory cortex in squirrel monkeys. *Journal of Neurophysiology, 74*, 722-732.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology, 50*, 5-17.

Schubotz, R. I., Friederici, A. D., & Yves von Cramon, D. (2000). Time Perception and Motor Timing: A Common Cortical and Subcortical Basis Revealed by fMRI. *NeuroImage, 11*, 1-12.

Sekuler, R., & Blake, R. (1994). *Perception* (3rd ed.). Singapore: McGraw-Hill, Inc.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*, 308-308.

Semjen, A., Schulze, H. H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research-Psychologische Forschung, 63*, 137-147.

Shailer, M. J., & Moore, B. C. J. (1983). Gap detection as a function of frequency, bandwidth, and level. *The Journal of the Acoustical Society of America, 74*, 467-473.

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry, 162*, 2384-2386.

Sherrick, C. E., & Cholewiak, R. W. (1986). Cutaneous Sensitivity. In K. R. Boff & L. Kaufmann & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 1). New York: John Wiley and Sons.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology, 11*, 505-509.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science 145*, 1328-1330.

Shore, D. I., Gray, K., Spry, E., & Spence, C. (2005). Spatial modulation of tactile temporal-order judgments. *Perception, 34*, 1251-1262.

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychological Science, 12*, 205-212.

Sinex, D. G. (1978). Cross-modality temporal resolution for auditory, vibrotactile and visual stimuli. *Journal of the Acoustical Society of America, 63*, S52.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception and Psychophysics, 69*, 673-686.

Smith, A., Taylor, E., Lidzba, K., & Rubia, K. (2003). A right hemispheric frontocerebellar network for time discrimination of several hundreds of milliseconds. *NeuroImage, 20*, 344-350.

Smith, W. F. (1933). The relative quickness of visual and auditory perception. *Journal of Experimental Psychology, 16*, 239-257.

Snell, R. S., & Lemp, M. A. (1998). *Clinical Anatomy of the Eye* (2nd ed.). Malden, MA: Blackwell Science, Inc.

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

Spence, C., Nicholls, M. E. R., & Driver, J. (2001a). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics, 63*, 330-336.

Spence, C., Shore, D. I., & Klein, R. M. (2001b). Multisensory prior entry. *Journal of Experimental Psychology-General, 130*, 799-832.

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt aftereffect. *Perception and Psychophysics, 62*, 1525-1533.

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. Cambridge, MA: MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*, 255-266.

Stein, B. E., Wallace, M. T., & Stanford, T. R. (2005). Brain Mechanisms for Synthesizing Information from Different Sensory Modalities. In E. B. Goldstein (Ed.), *Blackwell Handbook of Sensation & Perception*. Malden, MA: Blackwell Publishing.

Stelmach, L. B., & Herdman, C. M. (1991). Directed Attention and Perception of Temporal Order. *Journal of Experimental Psychology-Human Perception and Performance, 17*, 539-550.

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.

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. *Neuron, 51*, 651-659.

Stocker, A. A., & Simoncelli, E. P. (2006). Sensory Adaptation within a Bayesian Framework for Perception. In Y. Weiss & B. Scholkopf & J. Platt (Eds.), *Advances in Neural Information Processing Systems*. London: Springer.

Stohr, P. E., & Goldring, S. (1969). Origin of somatosensory evoked potentials in man. *Journal of Neurosurgery, 31*, 117-127.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M., & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*, 31-38.

Stone, J. V., Kerrigan, I. S., & Porrill, J. (in press). Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings of the Royal Society B-Biological Sciences*.

Strasburger, H. (2001). Converting between measures of slope of the psychometric function. *Perception and Psychophysics, 63*, 1348-1355.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception - Implicit estimation of sound-arrival time. *Nature, 421*, 911-911.

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience, 1*, 183-184.

Symmes, D., Chapman, L. F., & Halstead, W. C. (1955). The Fusion of Intermittent White Noise. *The Journal of the Acoustical Society of America, 27*, 470-473.

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition, 17*, 219-239.

Takahashi, K., Saiki, J., & Watanabe, K. (2008). Realignment of temporal simultaneity between vision and touch. *NeuroReport, 19*, 319-322.

Tanner, T. A., Patton, R. M., & Atkinson, R. C. (1965). Intermodality judgments of signal duration. *Psychonomic Science, 2*, 271-272.

Tappe, T., Niepel, M., & Neumann, O. (1994). A Dissociation between Reaction-Time to Sinusoidal Gratings and Temporal-Order Judgment. *Perception, 23*, 335-347.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *Journal of the Acoustical Society of America, 41*, 782-787.

Teatini, G., Farne, M., Verzella, F., & Berruecos, P. (1976). Perception of temporal order: visual and auditory stimuli. *Giornale Italiano Di Psicologia, 3*, 157-164.

Tepas, D. I., & Armington, J. C. (1962). Properties of visual evoked potentials. *Vision Research, 2*, 449-461.

Terao, M., Watanabe, J., Yagi, A., & Nishida, S. (2008). Reduction of stimulus visibility compresses apparent time intervals. *Nature Neuroscience, 11*, 541-542.

Tesche, C. (2000). Evidence for somatosensory evoked responses in human temporal lobe. *NeuroReport, 11*, 2655-2658.

Thomas, G. J. (1941). Experimental study of the influence of vision on sound localisation. *Journal of Experimental Psychology, 28*, 167-177.

Thurm, U. (1983). Mechano-electric transduction. In W. Hoppe & W. Lohmann & H. Markl & H. Ziegler (Eds.), *Biophysics*. Berlin: Springer-Verlag.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. New York: Macmillan.

Todd, J. W. (1912). *Reaction to multiple stimuli*. New York: The Science Press.

Tregellas, J. R., Davalos, D. B., & Rojas, D. C. (2006). Effect of task difficulty on the functional anatomy of temporal processing. *NeuroImage, 32*, 307-315.

Treutwein, B. (1995). Adaptive Psychophysical Procedures. *Vision Research, 35*, 2503-2522.

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research, 149*, 439-446.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics, 66*, 1171-1189.

Turatto, M., Benso, F., Galfano, G., & Umilta, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology-Human Perception and Performance, 28*, 628-639.

Ulrich, R., Nitschke, J., & Rammsayer, T. H. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research-Psychologische Forschung, 70*, 77-87.

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*, 915-928.

Uttal, W. R. (1959). A comparison of neural and psychological responses in the somesthetic system. *Journal of Comparative Physiology and Psychology, 52*, 485-490.

van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex, 17*, 962-974.

van Doren, C. L., Gescheider, G. A., & Verrillo, R. T. (1990). Vibrotactile temporal gap detection as a function of age. *Journal of the Acoustical Society of America, 87*, 2201-2206.

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*, 955-968.

van Erp, J. B. F., & Werkhoven, P. J. (2004). Vibro-tactile and visual asynchronies: Sensitivity and consistency. *Perception, 33*, 103-111.

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of Subjective Time Perception Within and Across Senses. *PLoS ONE, 3*, e1437.

Vatakis, A., Bayliss, L., Zampini, M., & Spence, C. (2007). The influence of synchronous audiovisual distractors on audiovisual temporal order judgments. *Perception and Psychophysics, 69*, 298-309.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008). Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Experimental Brain Research, 185*, 521-529.

Verrillo, R. T. (1963). Effect of Contactor Area on the Vibrotactile Threshold. *Journal of the Acoustical Society of America, 35*, 1962-1966.

Verrillo, R. T. (1982). Effects of aging on the suprathreshold responses to vibration. *Perception and Psychophysics, 32*, 61-68.

Vibell, J., Klinge, C., Zampini, M., Spence, C., & Nobre, A. C. (2007). Temporal order is coded temporally in the brain: Early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *Journal of Cognitive Neuroscience, 19*, 109-120.

Vogels, I. (2004). Detection of temporal delays in visual-haptic interfaces. *Human Factors, 46*, 118-134.

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*, 32-35.

Walker, J. T., & Scott, K. J. (1981). Auditory-Visual Conflicts in the Perceived Duration of Lights, Tones, and Gaps. *Journal of Experimental Psychology-Human Perception and Performance, 7*, 1327-1339.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence. *BMC Neuroscience, 9*.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology, 48*, 242-259.

Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics, 62*, 1258-1264.

Waszak, F., Wascher, E., Keller, P., Koch, I., Aschersleben, G., Rosenbaum, D. A., & Prinz, W. (2005). Intention-based and stimulus-based mechanisms in action selection. *Experimental Brain Research, 162*, 346-356.

Watanabe, J., Amemiya, T., Nishida, S., & Johnston, A. (2008). Independent local adaptation of tactile frequency and duration. *Perception 37*, S87.

Watson, A. B., & Fitzhugh, A. (1990). The method of constant stimuli is inefficient. *Perception and Psychophysics, 47*, 87-91.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*, 113-120.

Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why "sounds are judged longer than lights": Application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology, 51*, 97-120.

Wearden, J. H., Todd, N. P. M., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *Quarterly Journal of Experimental Psychology, 59*, 1709-1724.

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., Hallett, M., & Rauschecker, J. P. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *Journal of Neuroscience, 20*, 2664-2672.

Weeks, R. A., Honda, M., Catalan, M. J., & Hallett, M. (2001). Comparison of Auditory, Somatosensory, and Visually Instructed and Internally Generated Finger Movements: A PET Study. *NeuroImage, 14*, 219-230.

Weisenberger, J. M. (2005). Cutaneous Perception. In E. B. Goldstein (Ed.), *Blackwell Handbook of Sensation and Perception*. Marden, MA: Blackwell Publishing Ltd.

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics, 39*, 294-300.

Welch, R. B., & Warren, D. H. (1980). Immediate Perceptual Response to Intersensory Discrepancy. *Psychological Bulletin, 88*, 638-667.

Welch, R. B., & Warren, D. H. (1986). Intersensory Interactions. In K. R. Boff & L. Kaufmann & P. James (Eds.), *Handbook of Perception and Human Performance*. New York: Wiley.

Westheimer, G. (1999). Discrimination of short time intervals by the human observer. *Experimental Brain Research, 129*, 121-126.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America, 85*, 858-867.

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *Journal of the Acoustical Society of America, 85*, 868-878.

Wohlschlager, A., Engbert, K., & Haggard, P. (2003). Intentionality as a constituting condition for the own self-and other selves. *Consciousness and Cognition, 12*, 708-716.

Wohlschlager, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self- and other-generated actions. *Psychological Science, 14*, 586-591.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., & Lederman, S. J. (2006). *Sensation & Perception*. Sunderland, MA: Sinauer Associates.

Woolsey, T. A., Hanaway, J., & Gado, M. H. (2003). *The Brain Atlas: A Visual Guide to the Human Nervous System*. Hoboken, New Jersey: John Wiley.

Xuan, B., Zhang, D., He, S., & Chen, X. C. (2007). Larger stimuli are judged to last longer. *Journal of Vision, 7(10):2*, 1-5.

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature, 414*, 302-305.

Yarrow, K., & Rothwell, J. C. (2003). Manual chronostasis: Tactile perception precedes physical contact. *Current Biology, 13*, 1134-1139.

Zampini, M., Brown, T., Shore, D. I., Maravita, A., Roder, B., & Spence, C. (2005a). Audiotactile temporal order judgments. *Acta Psychologica, 118*, 277-291.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005b). Audio-visual simultaneity judgments. *Perception & Psychophysics, 67*, 531-544.

Zampini, M., Shore, D. I., & Spence, C. (2003a). Audiovisual temporal order judgments. *Experimental Brain Research, 152*, 198-210.

Zampini, M., Shore, D. I., & Spence, C. (2003b). Multisensory temporal order judgments: the role of hemispheric redundancy. *International Journal of Psychophysiology, 50*, 165-180.

Zampini, M., Shore, D. I., & Spence, C. (2005c). Audiovisual prior entry. *Neuroscience Letters, 381*, 217-222.

Zhou, W., & Chen, L. (1994). Modulation of motion aftereffects (MAE) by engagement of attention in an auditory task. *Investigative Ophthalmology and Visual Science, 35(supplement)*, 1838.