Multisensory integration of redundant and complementary cues

Jessica Hartcher-O'Brien

Linacre College, University of Oxford

Declaration

I declare that this work has not been submitted previously as an exercise for a degree at this or any other university, and that it is entirely my own work.

Jessica Hartcher-O'Brien

Contents

CHAPTER 2: MULTISENSORY INTERACTIONS IN SPACE: EXTENDING THE COLAVITA VISUAL DOMINANCE EFFECT TO TOUCH 29

CHAPTER 3: EXTENDING THE VISUAL DOMINANCE OVER TOUCH FOR INPUT OFF THE BODY 71

CHAPTER 5: INTERRUPTING EXPOSURE TO ASYNCHRONY LEADS TO THE SEPARATION OF THE SENSORY INPUTS, NOT RECALIBRATION 109

List of figures

List of tables

List of abbreviations

AV	Auditory leading asynchrony
JND	Just noticeable difference
PSS	Point of subjective simultaneity
PSE	Point of subjective equality
RT	Reaction time
SC	Superior colliculus
SJ	Simultaneity judgment
TOJ	Temporal order judgment
WF	Weber fraction
VA	Vision leading asynchrony
SOA	Stimulus onset asynchrony
ANOVA	Analysis of variance
RM	Repeated measures

Publications from present work

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (under revision). The duration of uncertain times: audiovisual information about intervals is integrated in a statistically optimal fashion.

Hartcher-O'Brien, J., Levitan, C. A., & Spence, C. (2010). Extending the visual dominance of touch for input off the body. *Brain Research,* **1362**, 48–55.

Navarra, J., **Hartcher-O'Brien, J**., Piaza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of a sound. *Proceedings of the National Academy of Sciences, USA*, **106**, 9169–9173.

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.

Other publications

Conrad, C., Kleiner, M., **Hartcher-O'Brien, J.**, Bartels, A., Bülthoff, H. H., & Noppeney, U. (under revision). Audio-visual interactions in binocular rivalry using the Shepard illusion in the auditory and visual domain.

Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *Journal of Experimental Psychology: Human Perception & Performance*, **37**, 1383–1395.

Occelli, V., **Hartcher-O'Brien, J.**, Spence, C., & Zampini, M. (2010). Assessing the audiotactile Colavita effect in near and rear space. *Experimental Brain Res*earch, **203**, 517–532.

Auvray, M., Gallace, A., **Hartcher-O'Brien, J**., Hong, T., & Spence, C. (2008). Tactile and visual distractors induced change blindness for tactile stimuli presented on the fingertips. *Brain Research*, **1213**, 111–119.

Abstract

Multisensory integration of redundant and complementary cues

Jessica Hartcher-O'Brien

Linacre College, University of Oxford
DPhil, Trinity term 2012

During multisensory integration, information from distinct sensory systems that refers to the same physical event is combined. For example, the sound and image that an individual generates as s/he interacts with the world, will provide the nervous system with multiple cues which can be integrated to estimate the individual's position in the environment. However, the information that is perceived through different sensory pathways/systems can be qualitatively different. The information can be redundant and describe the same property of an event in a common reference frame (i.e., the image and sound referring to the individual's location), or it can be complementary. Combining complementary information can be advantageous in that it extends the range and richness of the information available to the nervous system, but can also be superfluous and unnecessary to the task at hand – i.e. olfactory cues about the individuals perfume can increase the richness of the representation but not necessarily aid in localisation. Over the last century or so, a large body of research has focused on different aspects of multisensory interactions at both the behavioural and neural levels. It is currently unclear whether the mechanisms underlying multisensory interactions for both type of cue are similar or not. Moreover, the evidence for differences in behavioural outcome, dependent on the nature of the cue, is growing. Such cue property effects possibly reflect a processing heuristic for more efficient parsing of the vast amount of sensory information available to the nervous system at any one time.

The present thesis assesses the effects of cue properties (i.e., redundant or complementary) on multisensory processing and reports a series of experiments demonstrating that the nature of the cue, defined by the task of the observer, influences whether the cues compete for representation as a result of interacting, or whether instead multisensory information produces an optimal increase in reliability of the event estimate. Moreover, a bridging series of experiments demonstrate the key role of redundancy in inferring that two signals have a common physical cause and should be integrated, despite conflict in the cues. The experiments provide insights into the different strategies adopted by the nervous system and some tentative evidence for possible, distinct underlying mechanisms.

Extended abstract

This thesis explores how complementary and redundant cues are used by the nervous system in the processing and integration of multisensory information that specifies spatial and temporal properties of the environment. The influence and effects of these two cue types are assessed and empirically tested.

In the initial study, described in Chapter 2, observers were presented with a visual flash, a tactile vibration to the fingertip, or both cues simultaneously. Their task was to respond as quickly as possible to the visual flash or vibration or, when presented simultaneously, to both. Responses to the tactile pulse and visual flash were highly accurate. Responses to the two cues presented simultaneous on the other hand, were not: Observers failed to respond to the tactile pulse cue on a significant number of these multisensory trials. This visual dominance for detecting the flash and inhibiting the tactile pulse was stronger for stimuli at threshold (Experiment 4). When the probability of multisensory cues was increased, the visual dominance decreased (Experiment 2). In this study, the visual flash and tactile vibration presented simultaneously were complementary as they did not specify the same object property, but were discriminated as two co-occurring inputs. Inhibition of the vibration cue was observed as a result.

In the experiment described in Chapter 3, the question was whether this visual-tactile interaction was stronger when occurring in personal, rather than extrapersonal space.

Observers performed the same speeded detection task to flashes and tactile pulses. However, 33% of visual flashes were either presented to one of three locations: the observers fingertip, a distant location in extrapersonal space, or to a rubber hand at the same distant location in extrapersonal space. Unisensory visual responses to the simultaneously presented flashes and tactile pulses occurred when the visual flash was presented on the observer's body, or on the rubber hand despite it being located in extrapersonal space. The visual dominance effect did not transfer to extrapersonal space when the rubber hand was not present. These results therefore suggest that the competition between the complementary flash and tactile pulse is strongest for events on a body surface.

In Chapters 2 and 3 the inhibition of touch effect occurred in the context of speeded responses to an unpredictable temporal sequence of unisensory, visual flashes, tactile pulses and flash-pulse cue pairs. The nature of the temporal sequence may play a role given that when the probability of multisensory trials was increased, the effect decreased. In order to explicitly test the influence of *multisensory* temporal sequences and the statistics of these on subsequent speeded detection responses, Chapters, 4, 5, and 6 explored the phenomenon of temporal recalibration for audiovisual cues.

In the study discussed in Chapter 4, the influence of short-term temporal sequences on responses was explicitly tested. Observers were exposed to a continuous train of redundant but slightly discrepant beeps and flashes. After this exposure period, they were required to respond as rapidly as possible whenever they detected unisensory beeps or flashes of light.

Response latencies (RTs) to these unisensory signals changed as a result of the type of discrepancy presented in the exposure phase. If the beeps preceded the flashes of light then the consequent unisensory auditory RTs were slower than when the beep lagged behind the flash in the exposure period. This change in auditory RTs is consistent with the nervous system recalibrating to the temporal discrepancy between the beep and flash. This finding demonstrates the importance of the statistics of the environment, established via short-term temporal sequences.

In the study reported in Chapter 5, observers were again exposed to redundant, discrepant beep and flash sequences. But this time exposure was intermittent. After 40 seconds of being presented with the discrepant beeps and flashes, observers were provided with unisensory beeps and flashes and required to detect and respond as quickly as possible to these unisensory signals. The change in the RTs to the unisensory beeps and flashes was measured over time. However, neither the auditory nor the visual RTs shifted after the intermittent exposure to the discrepant beeps and flashes. The lack of observed recalibration highlights an important pre-requisite for temporal recalibration to occur: exposure to the discrepancy needs to be continuous. Otherwise the cues, though redundant, are segregated and treated as arising from two distinct events.

The work reported in Chapter 6 explored a slightly different aspect of recalibration: perceived simultaneity of the beep and flash after exposure. Following continuous and prolonged exposure to a beep of one auditory frequency and a flash occurring at a slightly different time, observers made SJ to single beep-flash pairs. The beep-flash pair either contained the exposed beep frequency or a beep frequency not presented

before. Perceived simultaneity only changed for the adapted auditory frequency. The un-adapted frequency signal was not modulated. As such the nervous system may have inferred that this cue arose from an event separate to that to have undergone recalibration.

From these temporal recalibration studies it can be concluded that when the nervous system is presented with redundant, discrepant beeps and flashes it shifts the timing of the beep if it can infer with reasonable likelihood, that the beep and flash arose from a single event. This response strategy is inline with a unisensory adjustment process (see Di Luca, et al., 2009).

The final study presented in Chapter 7 explored how the nervous system estimates the duration of a redundant sound and flash event. Observers were presented with either two intervals of sound, two intervals defined by flashes, or two intervals of simultaneously presented sounds and flashes. The observers had to indicate if the first or second interval (irrespective of the type of signal, sound, light flash or both) was longer. When they judged the combined sound and flash intervals their estimates were more precise than either the light flash or sound alone. That is, when the nervous system was provided with redundant cues it had an advantage that reduced uncertainty in its interactions with the environment. This finding highlights the benefit observed for multisensory integration.

The findings presented in this thesis cover a range of perceptual challenges faced by the nervous system in our multisensory world. Figure 7.1 outlines an initial suggestion of the mechanisms hypothesised to underlie these processes for both redundant and complementary cues. In the following sections the results summarised above will be discussed in terms of cue properties and potential mechanisms.

Chapter 1: Introduction

The human nervous system is continuously presented with multisensory information coded by multiple sensory channels. Yet how this information is combined in order to form a coherent representation of the environment is, in itself, a puzzle. This thesis provides insights into how redundant and complementary multisensory cues are processed and integrated across both spatial and temporal attributes of the environment. A number of different paradigms and approaches are implemented in order to explore the mechanisms underlying these processes. Solving the puzzle of how multisensory cues are appropriately combined is elaborated in the sections that follow.

1.1 Multisensory Integration across space and time

Imagine you are in a crowded room and you hear a familiar voice. A natural response is to look in the general direction of the source of the sound. However, in order for you to identify the speaker correctly, being able to detect and observe spatiotemporally coincident voice and lip movements will increase your chances of identifying the right individual.

The individual addressing you provides a vast amount of continuous stimulation to your nervous system. Our everyday experiences present the nervous system with even more information that is processed by multiple sensory channels, tuned to specific forms of energy. These different kinds of energy, in turn, code for different aspects of the environment. For us to efficiently interact with the events and objects around us, such as the speaker, the information from different sources must eventually converge into a coherent and unambiguous multisensory perception of the relevant environmental events. The capacity of the central nervous system to combine inputs across the senses can lead to marked improvements in the detection, localization, and discrimination of external stimuli and to faster reactions to those stimuli (e.g., Frassinetti, Bolognini, & Ladavas, 2002; King, 2004; King & Calvert, 2001; Lippert, Logothetis, & Kayser, 2007; Stein London, Wilkinson, & Price, 1996).

Traditionally, multisensory integration has been operationalised as an enhancement in neural responses or an increase in a perceptual estimate's precision (see Ernst & Bülthoff, 2004; Stein et al., 1987, 1993). This can be traced back to the physiological school emerging from the work of Stein and Meredith (1993). The increase in precision, not necessarily subserved by the same mechanisms as the enhancement in the neural sense, provides an easy litmus test for whether combination leads to integration (see Ernst & Di Luca, 2011). However, one particularly puzzling phenomenon does not support this operationalisation: why multiple sensory inputs would result in the inhibition of one of the components (Colavita, 1974). Here multisensory integration is defined as the process by which the similar or dissimilar perceptual features, carried by distinct sensory modalities, are bound together across both spatial and temporal dimensions, in order to form a common coherent representation of the environmental events. However, the criteria and the mechanisms underlying the integration of the multiple sensory cues remain poorly understood.

1.1.1 Original principles of multisensory integration

Two primary branches of multisensory perception have developed in order to understand how multisensory integration is achieved by the nervous system: the one based on physiology and the other based more on the perceptual, computational school of thought (see Stein, 2010). Meredith and his colleagues (Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1986) provided ground-breaking insights into the problem of multisensory integration from the physiological perspective at the neuronal level. They investigated the response characteristics of superior colliculus (SC) neurons of the anaesthetised cat to redundant multisensory inputs (Meredith & Stein, 1993; see also King & Palmer, 1988). With single cell recordings, they measured the changes in the neural response for multiple sensory inputs. They systematically manipulated the temporal, spatial and intensity relationships between auditory and visual information. From the response characteristics in their neural data they were able to distil three principles by which multisensory integration is governed:

i. <u>The spatial rule:</u> When multiple sensory inputs are spatially coincident, the strength of (multisensory) neural responses is maximal. The probability that the neurons will respond is highest when the signals arise from the same location. The response magnitude decreases with increasing physical distance between the two unisensory inputs.

ii. <u>The temporal rule:</u> multisensory signals occurring simultaneously are more likely to elicit a response from multisensory neurons in the SC. Increasing the temporal discrepancy/delay between the two unisensory inputs decreases the strength of the neural response.

iii. <u>The principle of inverse effectiveness:</u> Strong multisensory integration is more likely to occur if the component unisensory signals presented alone evoke only a weak response from a neuron.

These principles define the circumstances under which integration of multiple sensory inputs, at the neuronal level of the SC, should occur. Although this pioneering research was conducted on individual SC neurons in the cat, the principles themselves have guided the field of multisensory research at both the physiological and behavioural levels since (see Alais, Newell, & Mamassian, 2010).

Most organisms are able to integrate information across their sensors and use spatial proximity as well as temporal coincidence in order to infer that multiple sensory signals arose from a common cause and should therefore be integrated (Körding, Beierholm, Ma, Quartz, Tenenbaum, & Shams, 2007; Lewald & Guski, 2004). The coherent unambiguous percept that arises from integration can then be used to select and control appropriate responses to the incoming information (Lalanne & Lorenceau, 2004). Approximate, as opposed to identical, physical location can be used by the nervous system as suggested by the first neural response mechanism of multisensory integration described above. In order to decide whether to integrate redundant signals or not, the nervous system has been shown to adopt a principle of 'near enough is good enough'. Yet in coding for differences or similarity in position or timing, the senses themselves are inherently noisy. If the multisensory signals only vary in the amount of noise they contain, or their latency, it is important to consider what consequence this may have for perception and action. One possibility is that the nervous system has a way of computing and compensating for such factors. This idea is explored further in Section 1.1.2. To address this issue, it is important to consider how the different sensory channels arrive at an estimate of the position or timing of an event. Do differences in these estimates only arise from noise? How does the nervous system deal with discrepancies caused by noise versus discrepancies caused by physical differences across the stimuli? The ability of the nervous system to correctly bind information has been characterized as the correspondence, or binding, problem (Marr, 1982). Although correspondence across the sensory channels has been often found to increase the precision of the bimodal estimate (e.g., for location, Alais & Burr, 2004; for size, Ernst & Banks, 2002; for the felt position of a limb, van Beers et al., 1999), this is not always the case (e.g., Brayanov & Smith, 2010; Hartcher-O'Brien & Alais, 2011).

The second principle, that of temporal coincidence, is one of the strongest cues used to signal that multiple sensory inputs should be integrated (Burr, Silva, Cicchini, Banks, & Morrone, 2009b; Calvert, Spence, & Stein, 2004; de Gelder, & Bertelson, 2003; Sekular, Sekular, & Lau, 1997; Spence & Squire, 2003; see Keetels & Vroomen, 2010, for a review). This cue allows the nervous system to estimate (with a higher probability than when it is absent) that the sensory signals arose from the same environmental event. Relatively large physical asynchronies (in the order of hundreds of milliseconds) can be tolerated by the nervous system and signals will still be integrated. On the basis of prior behavioural research, we know that the flexibility of the system depends on the type of signals being integrated (e.g., Keetels & Vroomen, 2010; Vatakis et al., 2007, 2008; Vroomen & Keetels, 2010) and the latencies of the sensory systems involved (King, 2005): Impoverished signals like beeps and flashes are unlikely to be bound when the temporal discrepancy is greater than 150ms (Lewald, Ehrenstein, & Guski, 2001; Navarra, Vatakis, Zampini, Soto-Faraco, Humphreys, & Spence, 2005). For rich contextual cue signals such as speech, the window of multisensory integration is somewhat broader – i.e., 250-300ms depending on the order of stimulus presentation (van Wassenhove, Grant, & Pöppel, 2005). This window provides insight into the tolerance of the sensory system and the constraints of sensory processing that influence multisensory integration. Other temporal properties of the environment can be utilised by the nervous system in order to infer a common cause and integrate multiple cues. For example, a common duration, i.e., things remaining for the same amount of time on the sensory receptors, also provides valuable information to the nervous system that what is seen and heard arose from the same underlying object. Yet the question of mechanisms for multisensory integration of redundant signals for temporal aspects of the environment is far more contentious and problematic than for spatial properties. These include estimations of multisensory elapsed time (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011; Lewis & Miall, 2003; van Wassenhove, Buonomano, Shimojo, & Shams, 2008) and temporal recalibration (see Vroomen & Keetels, 2010, for a review). The principles of multisensory integration (Stein & Meredith, 1993) can guide the nervous system in its assignment of a common cause for the multiple sensory cues.

There are many different forms of multisensory interactions. The ventriloquist effect provides well-known example and is often used to investigate possible mechanisms of multisensory integration (see Alais & Burr, 2004; Howard & Templeton, 1966). The effect arises when slightly discrepant audio and visual signals are perceived as arising from the same location. One of the sensory components is mislocalised by the nervous system in the direction of the more reliable signal (e.g., Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003). This mislocalisation has been explained mathematically as a weighted average of the sensory components. This mathematical model has also been applied to many other behavioural phenomena such as, temporal adaptation effects (Di Luca, Machulla, & Ernst, 2009; see also Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009), the two-flash illusion (Andersen, Tiippana, & Sams, 2005; Bresciani, Dammeier, & Ernst, 2006; Shams, Ma, & Beierholm, 2005; Wozny, Beierholm, & Shams, 2008), cross-modal effects on rate perception (Roach, Heron, & McGraw, 2006), and visual-vestibular interactions (Angelaki et al., 2009; Butler, Smith, Campos, & Bülthoff, 2010). This approach is known as part of the Bayesian framework discussed further in Section 1.1.2. It has recently emerged as *the* model of cue integration for multisensory perception (see Ernst & Di Luca, 2011, for a review). Traditionally, phenomena such as the ventriloquist effect (Bertelson, 1996; Howard & Templeton, 1966), have been described in terms of the "modality-appropriateness hypothesis" (MAH) (Welch & Warren, 1986). In the Bayesian approach, cues are weighted online according to the variability contained in each cue estimate: the result of a flexible nervous system. MAH on the other hand posits a more inflexible hardwired system such that the cues are integrated according the acuity of the sensory channel transmitting the information. In light of the extraordinary plasticity of the nervous system (e.g., Bach-y-Rita, 1967; King, Parsons, & Moore, 2000) this is unlikely. Moreover, the Bayesian approach (as discussed in this thesis) provides one major advance not captured in the MAH: the ability to predict behaviour.

1.1.2 Models & mechanisms of multisensory integration – a noisy issue

Human perception is inherently multisensory but also inherently noisy. In most instances multiple sources of information redundantly specify an environment property i.e., the location of the person addressing you. Redundancy allows the system to reduce the amount of uncertainty in its estimate of, for example, the location of the speaker. Combining information across the senses is essential as no information processing system – biological or technical – has the computational power to provide a precise and accurate estimate in all circumstances. Furthermore, all information-processing systems (including the nervous system) are perturbed by noise at all levels of processing. However, the nervous system can utilise common spatial and temporal properties to infer redundancy in the incoming information.

The process of combining information across the senses to form a coherent percept of the environment is complicated by noise at every level of processing, neuronal and otherwise (Ernst & Bülthoff, 2004). This suggests that perceptual estimates of environmental properties should therefore be unreliable and ambiguous. There are, however, several strategies that the nervous system may use in order to minimize noise in its final estimate of the input. One common approach exploits the interplay between vision and audition to estimate, for example, the position of objects in space – sensory interactions for object size, orientation etc. have also been investigated (Ernst & Banks, 2002). Combining sensory cues across the channels by a process of averaging allows the nervous system to reduce the effects of noise.

Using an averaging strategy, on the other hand, is most beneficial if the noise sources associated with the redundant estimates are independent of one another (Clarke &

Yuille, 1990). It must be noted that weighted averaging is also beneficial with correlated noise sources, as long as the noise sources are not totally correlated. However, the independence of noise sources for multisensory inputs will be discussed further. Coming back to the person calling your name. The auditory and visual information come from the two channels independently, either obtained from photons on the retina or from differences in the time of arrival (or intensity) differences at the two ears. Thus both channels redundantly specify information about the speaker's location. Studies of spatial properties of the environment have shown that when integrating visual and auditory estimates corrupted by the same amount of noise, observers showed a √2 improvement in discrimination performance (e.g., Alais & Burr, 2004; Ernst & Banks, 2002). This is the optimal solution as it denotes the maximal possible reduction in variance. Every time you estimate where the speaker is your estimate will vary slightly, forming a probability density distribution of responses (the likelihood function). This probability function will also occur for the visual position estimates. Each estimate of spatial location (the visual S_V & the auditory S_A) is corrupted by noise (represented by σ_V & σ_A, respectively) as discussed above.

According to the Maximum Likelihood integration model (Clarke & Yuille, 1990), the integrated estimate S_{VA} is a weighted average of the individual sensory signals with weights w_i that sum up to one (the index i refers to the individual modalities V or A):

$$S_{VA} = \sum_i w_i S_i \qquad \text{where} \qquad \sum_i w_i = 1 \tag{1.1}$$

In the terms of this model, to behave in an optimal manner, i.e., with the least amount of variance in your estimate, the chosen weights are taken to be proportional to the reliability r (inverse variance, Eqn. 2) of any given signal. In other words, if the visual modality provides more reliable information in a given situation, such a signal would be given a higher weight in the averaging process.

$$w_j = \frac{r_j}{\sum_{i=1\ldots j\ldots N} r_i} \quad \text{with} \quad r_i = \frac{1}{\sigma_i^2} \tag{1.2}$$

The variance of the combined estimate will be less than that of either of the individual estimates feeding into the averaging process. That is, the reliability improves when information is combined. According to this integration principle the reliability of the combined estimate r is the sum of the reliabilities of the individual estimates:

$$r = \sum_i r_i \tag{1.3}$$

Using the predictions of this integration scheme, multisensory integration can be described as the result of an optimal process (Clarke & Yuille, 1990). For spatial estimates, averaging according to this integration scheme seems to be a general strategy utilised by the perceptual system (e.g., Alais & Burr, 2004; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003). The elegance of this model as opposed to the MAH (Welch & Warren, 1980) is that it does not presuppose one modality to be more precise or appropriate than another but instead adopts a flexible approach which simply combines the spatial location estimates from each modality using a weighting scheme that weights each signal based on how reliable it is (see Ernst & Bülthoff, 2004, for a review). But for multisensory integration in the time domain it has not been so straightforward to reach an understanding of the integration strategy adopted by the nervous system.

It is possible to use time to identify the temporal relationship between events and to anticipate actions. The degree to which the nervous system can exploit the temporal contingencies depends on the amount of variability or uncertainty contained in the events themselves but also the amount of change in the context over time. The question of how temporal context can influence estimates of the current state of the environment can be explored both implicitly and explicitly. More generally, knowledge about the distribution of events prior to the current estimate or response can facilitate efficient interactions with the environment. As such, temporal properties such as the history of previous information – i.e., a sequence – provide additional cues that the multisensory information arises from a single source and should therefore be integrated (see Burge, Ernst, & Banks, 2008; Parise, Ernst, & Spence, 2011). Recent research suggests that the statistics of the signals themselves play an important role in the magnitude of the multisensory interactions (Ernst & Di Luca, 2011; Ma, Beck, & Pouget, 2008; Parise et al., 2011). The role of the history of the sensory signals is discussed further in Section 1.3.

1.2 Redundant and complementary cues

Imagine again that you are in the crowded room referred to at the beginning of this chapter. You hear a familiar voice and this time they are calling your name. You turn to see who it is. You now see several people in the general direction the voice came from. Many are talking. Which one called your name? You hear the voice again and now the sound seems to come from approximately straight ahead. In your field of view, there are a number of possible individuals who may have called your name. Finally, you see one whose lips move as you hear your name once more. Sound and sight have come together and you identify the speaker as your good friend. How does the nervous system find the appropriate auditory–visual correspondence to determine that a sound and sight have come from the same source?

Information that is perceived through different sensory pathways differs in qualitative terms. That is, the senses can provide either redundant or complementary (non-redundant) information about the friend calling your name. Redundant sensory signals provide information about the same object property (e.g., location) and are represented in a common reference frame by the same units of measure. Continuing with the example of localizing and identifying your friend, vision and audition provide redundant information about the individual's location. Vision and smell, on the other hand, provide complementary information about the identity of the person.

In general, we benefit from combining multiple redundant sources of information. Having a redundant nervous system allows the system to optimise its processing of, and interaction with incoming information. For example, redundancy in joint degrees of freedom means that the nervous system is flexible and can optimise the joint configuration for task and goal. Yet not all multisensory cues in the environment are redundant. Correlated cues provide additional information, which allows the nervous system to construct a more complex, wholesome representation of the environment. Returning to the example of the person calling your name, if the visual and auditory information arise from the same location and have the same temporal structure, information is redundant and will be integrated. Both types of cue enrich the possible representation of the scene at hand. However complementary information can be unnecessary when trying to localise your friend. When this is the case, the information tends to compete for representation (e.g., Simonotto, Riani, Seife, Roberts, Twitty, &

Moss, 1997; Watson & Humphreys, 1997). As such these cues are candidates for suppression when task demands are high and processing resources are limited (Desimone & Duncan, 1995).

Depending on the nature of information, different strategies are used to combine sensory information derived from different sources. While the MLE integration model described in Section 1.1.2 requires redundant sensory estimates (Ernst & Bülthoff, 2004), other forms of interaction can occur among complementary cues. For example, when searching out your friend in the crowd, the characteristic perfume smell is a complementary cue that is likely to influence the interpretation of the individual as your friend versus as someone else. In this example, the identity can be disambiguated by complementary olfactory and auditory information gathered through exploring the scene for non-location related properties.

Combining complementary or redundant cues can alter the outcome of sensory integration (Ernst & Di Luca, 2011). But whether two cues are redundant or not, again depends on the observer's task: If the task is to localise the individual addressing you, then the auditory and visual information are redundant. If, on the other hand, the task is to respond as quickly as possible to any detected auditory or visual information in the current scene, then the visual and auditory information are no longer redundant.

Recent studies have highlighted the fact that complementary sources of information may not necessarily be combined to increase precision (see Kuschel, Di Luca, Buss, & Klatzky, 2010). These authors described how different attributes of a combined event provide complementary information, which is processed according to a different rule than redundant sensory signals. An extreme example of the change in processing strategy adopted for complementary cues, has been described by Watson and Humphreys (1997) who provide evidence for the fact that the nervous system suppresses complementary features when the task demands require. In Table 1.1 the differences in processing and interaction of complementary and redundant information are outlined.

Table 1.1. The factors discussed in the current dissertation which exert an influence on the integration process of redundant and complementary cues. The final factors considered are the cue combination strategies adopted according to the existing literature.

Factors	Redundant cues	Complementary cues
Spatial coincidence	✓	✓
Temporal coincidence	✓	✓
Inverse effectiveness	✓	–
Shared object property	✓	X
Optimal integration	✓	X
Sensory competition	X	✓

1.2.1 Multisensory interactions during sensory competition

Perceptual studies have shown unequivocally that stimuli in different sensory modalities can powerfully interact under certain circumstances, to determine perception or behaviour. The exact nature of this interaction depends on many factors, for example, whether the sensory modalities provide redundant information about the environmental event and whether or not there is competition for response or decisional resources (see Ernst & Di Luca, 2011; Helbig & Ernst, 2008).

In the following section, the multisensory interactions discussed focus on a particular instance when the multiple sensory signals do not redundantly specify a single object-property and are also not integrated. Under these conditions, the behavioural response does not reduce the uncertainty of the combined estimate; rather the multisensory inputs produce a unisensory response. This behavioural outcome can be understood in light of the suggestion that complementary cues compete for representation and one component is inhibited due to competition between the incoming input streams (Watson & Humphreys, 1997). Why this information produces a winner-take-all response is discussed further in the following sections.

1.2.2 Complementary cues: Competition, representation, and response

As discussed throughout this chapter, most of our interactions with the world are multisensory. Returning again to the example of the friend calling your name, if the individual happens to be wearing their characteristic perfume this cue is complementary and it is possible that your nervous system will suppress it and focus on the visual information arising from the lip movements. According to the biased (or integrated) competition hypothesis (see Desimone, 1996; Desimone & Duncan, 1995), many brain systems (both sensory and motor) are fundamentally competitive in nature when processing complementary cue information. To process the visual, auditory and olfactory information emitted by your friend, perhaps representing the visual/auditory information alone is the fastest way of representing your target and initiating an appropriate response. It should be noted that this strategy will not necessarily produce the most reliable representation of the incoming information. As mentioned in the previous section, the task and goals of the observer influence the outcome of multisensory information processing. Moreover, with complementary cues under conditions of competition, increasing the representation of one source of information (i.e., visual information of your friend) to promote its presence in awareness may occur at a cost to concurrent representations (Desimone, 1996). Desimone and Duncan proposed that at a neural level, the representation of different cues (or sources of information) cannot simultaneously exist independently. Either they are integrated, or one cue will be inhibited. In the context of this thesis it should be noted that this process is not the predominant response to the environment but may be brought to the fore when the task demands require it – i.e., when making speeded detection responses.

When faced with an unpredictable temporal sequence of, auditory, visual, and audiovisual stimuli, detection and response are highly accurate. On a fraction of multisensory trials, however, the nervous system produces a unisensory visual response. This failure to detect/respond to the non-visual component is known as the Colavita effect (Colavita, 1974). It has been suggested that the effect is the result of sensory competition (see Spence, Parise, & Chen, 2012, for a review). Increasing the probability of multisensory events decreases the magnitude of the visual dominance effect (Koppen & Spence, 2007c). In fact, when multisensory trials occur with a probability of 90%, the effect is eliminated. Thus there appear to be a number of important factors producing the competition and subsequent suppression response: the complementary nature of the cues (Section 1.3; Watson & Humphreys, 1997); temporal context (Jazayeri & Shadlen, 2010); and the sensory pathways processing the information (Duncan & Desimone, 1995; Romei, Murray, Merabet, & Thut, 2007).

The nature of the cues presented to the nervous system and whether the temporal context of primarily unisensory information increases the competition between the visual and non-visual information on the multisensory trials are explored throughout this dissertation.

The example of the Colavita effect demonstrates that the result of the cue competition is a winner-take-all outcome for vision (Colavita, 1974; Hartcher-O'Brien et al., 2008; Hartcher-O'Brien, Levitan, & Spence, 2010; Koppen & Spence, 2007a & b; Sinnett et al., 2008). Moreover, there is increasing evidence for the visual representation of auditory signals early in the visual pathway for a variety of tasks (Giard & Peronnet, 1999; Romei et al., 2007). Multisensory cues may therefore produce a stronger visual representation given the concurrent activation of visual areas by auditory inputs. Even if this were not the case the same outcome could also be explained by a further claim of the biased competition model: The dominant neural representation suppresses the neural activation associated with the representation of the weaker stimulus (see Desimone, 1996). Indeed, neuroimaging studies have revealed that the presentation of a visual stimulus prior to an auditory signal results in inhibition in the response of the auditory cortex (Kayser & Logothetis, 2007; Kayser, Petkov, Augath, & Logothetis, 2007; though see also Inan et al., 2004).

Consistent with the assumptions of the biased competition hypothesis, signal reliability alone does not lead to the vision winner-takes-all outcome of the Colavita effect. What may also contribute is that concurrent stimuli in our environment continuously compete to drive neurons and gain representation at a level of awareness. In other words, particularly salient stimuli will have a competitive advantage not only defined here in terms of their reliability. Their representation neurally is probably increased via mechanisms such as attention (see McDonald Teder-Sälejärvi, Russo, & Hillyard, 2005). While the two possibilities are not mutually exclusive, they tackle the problem of multisensory interactions from the two different schools of thought outlined in Section 1.1. In Chapters 1 and 2 of this thesis, the robustness of the Colavita effect is tested with visual and tactile stimuli. One reason for using tactile stimuli is that touch is more likely than sound to elicit an immediate response (e.g., Gregory, 1976). In this case vision may not have a competitive advantage. The importance of the question of whether or not vision dominates over touch is supported by the fact that tactile stimuli have been claimed to be of great adaptive importance for humans, given that they are thought to signal direct contact between external stimuli and the body surface (i.e., a potentially dangerous situation requiring rapid reactions; see Gregory, 1967). On the basis of this suggestion, one might therefore expect that vision would not dominate over tactile perception. As such, in some circumstances – i.e., speeded detection – it may increase the representation of stimulation via this sensory channel.

1.2.2.1 Objective 1

The first objective of this thesis is to understand how complementary cues are processed when observers make speeded detection/discrimination responses to an unpredictable stream of unisensory and less frequent multisensory events. The Colavita paradigm will be used to explore why down-weighting of the non-visual component occurs. This effect is known to decrease somewhat with spatial separation of component multisensory cues (Koppen & Spence, 2007a; Hartcher-O'Brien et al., 2008; Hartcher-O'Brien et al., 2010). The effect, though occurring in an artificial environment, is used to explore how the nervous system down-weights or promotes certain sensory information and decides what to carry further to awareness. Why the visual component is more often carried on to awareness is currently unclear. On a moment-by-moment basis, the visual system is challenged with processing immense amounts of information entering the eyes (and additional information from other sensory channels) with only a small fraction reaching awareness. Other possible interpretations of the down-weighting of non-visual information comes from the findings that multisensory signals have been shown extensively to influence visual information processing (see Frassinetti, Maini, Benassi, Avanzi, Cantagalloc, & Farnè, 2010; Shams, 2010; Shams, Kamitani, & Shimojo, 2002), utilising the visual processing system to gain access to awareness (for a review, see Ghazanfar & Schroeder, 2006). As such it is hypothesised that vision is an unlikely candidate down-weighting in the context of complementary multisensory cues.

1.3 Explicit temporal properties

It has been posited that the key to whether sensory modalities or features of the sensory input compete is whether or not signals are redundant (Desimone & Duncan, 1995; Watson & Humphreys, 1997). Furthermore, the nervous system is able to adapt to the context or history of different cue combinations in order to more efficiently resolve competition and or conflicts in the multiple cues. Temporal recalibration (as discussed in the following sections) provides an ideal way tease apart the influences of redundancy and non-explicit temporal features and thus make it possible to predict whether the sensory components should integrate or compete.

1.3.1 Temporal recalibration: repetitive, redundant, discrepant cues

Multisensory stimuli originating from the same external event, such as the voice emanating from your friend can be perceived as asynchronous due to differential physical and neural delays (Alais & Carlile, 2005; King, 2005; Spence & Squire, 2003). Temporal recalibration is a phenomenon whereby the delay between multisensory signals is reduced due to prolonged exposure to redundant asynchronous audiovisual input. The measurable change in perceived simultaneity, resulting from the exposure to redundant discrepant signals, highlights the flexibility of the strategies of the nervous system to form a coherent representation of the world irrespective of small discrepancies (Fujisaki, Shimojo, Kashino, & Nishida, 2004). Recalibration tends to occur when two redundant "variables of stimulation yield discrepant information about a distal property" (Epstein 1975). An early example for audiovisual recalibration comes from the work of Radeau and Bertelson (1974). These authors were interested in exactly how a spatial discrepancy between audio and visual stimuli is resolved. Subjects had to indicate the perceived position of the sound prior to and after exposure to an audiovisual pair of stimuli with a spatial misalignment. Two distinct phenomena are involved during recalibration. The immediate effect of integration, that is, the two redundant signals are treated as though they are produced by the same event. Consequently, the locations from which the stimuli appear to originate are shifted toward one another. In Section 1.1.2 the resolution of a single instance of discrepancy was described as the result of a weighted average of the unisensory estimates. But in the Radeau and Bertelson (1974) experiment a second interesting effect is observed. Prolonged exposure to constant discrepancy leads to a marked shift in perception.

Signals that were previously perceived as arising from two separate locations were perceived to come from a single location after exposure.

To understand why the nervous system may realign discrepant multisensory inputs, consider a live television broadcast for which a reporter transmits from a remote location. Technical limitations in the television apparatus can cause the voice of the reporter to be desynchronized with the visual stream, usually making the visual signal lead. At first, the relative delay between the two streams of information creates the impression that the voice and the moving lips/face of the reporter appear to belong to two different individuals. Only after some time, by experiencing the correlation between the voice and lip movements, does the nervous system realise that there is a correspondence between the two streams of information and therefore works to resolve the temporal inconsistency. As you keep watching the television report – presented with the continuous discrepancy – the temporal offset becomes perceptually less noticeable. The audiovisual information associated with the reporter becomes unified into one coherent speech percept and it is then possible to focus on the content. This perceptual shift resulting from exposure to asynchrony implies a pure temporal shift, i.e., recalibration (see Arnold & Yarrow, 2011; Di Luca et al., 2009; Fujisaki et al., 2004; Hanson, Heron, & Whitaker, 2008b; Keetels & Vroomen, 2005; Navarra, Hartcher-O'Brien, et. al., 2009; Vatakis, Navarra, Soto-Faraco, & Spence, 2007; Vroomen, Keetels, de Gelder, & Bertelson, 2004). This is not the same result as observed with repeated exposure that results in other forms of adaptation (i.e., Foulkes & Miall, 2000; Hanson, Heron, & Whitaker, 2008a; Stetson, Montague, & Eagleman, 2006; Vatakis et al., 2008). Nor does it occur as a result of inhibition of one of the sensory components as occurs with a random sequence of complementary sensory cues. This distinction is important in terms of the mechanisms that may be responsible for the effect of temporal recalibration.

The basic paradigm for assessing this phenomenon is an experiment entailing an exposure phase and a test phase: An exposure phase to the discrepant pair of redundant signals and a test phase that depends on the method used to assess recalibration. In the test phase, each judgment is followed by a small re-exposure to the asynchrony in order to maintain a constant level of 'recalibration'. A standard method for quantifying the perceived temporal relation between stimuli is the TOJ task (Hirsh & Sherrick, 1961; Spence, Baddeley, Zampini, James, & Shore, 2003; Sternberg & Knoll, 1973; Zampini, Shore, & Spence, 2003a). The two stimuli are presented separated by a range of different asynchronies until the Point of Subjective Simultaneity (PSS) is obtained from the participant's responses. The PSS indicates the asynchrony at which stimuli must be presented to be perceived as being simultaneous. Alternatively, observers are asked to perform a simultaneity judgment (SJ) task, so that the asynchronies at which there is a majority of synchrony judgments define the "window of simultaneity" (sometimes called "window of integration"; Dixon & Spitz, 1980; Hirsh & Sherrick, 1961; Spence, Shore, & Klein, 2001; Zampini et al., 2003). A similar estimate can be achieved with a TOJ using the 75% boundaries. The middle of the simultaneity window is used as an estimate of the PSS (see Figure 1.1). If recalibration results from adjusting the timing or processing of individual but redundant stimuli (Di Luca et al., 2009; Navarra et al., 2009), then recalibration should also produce a change in response latency, i.e., reaction times to auditory and visual stimuli.

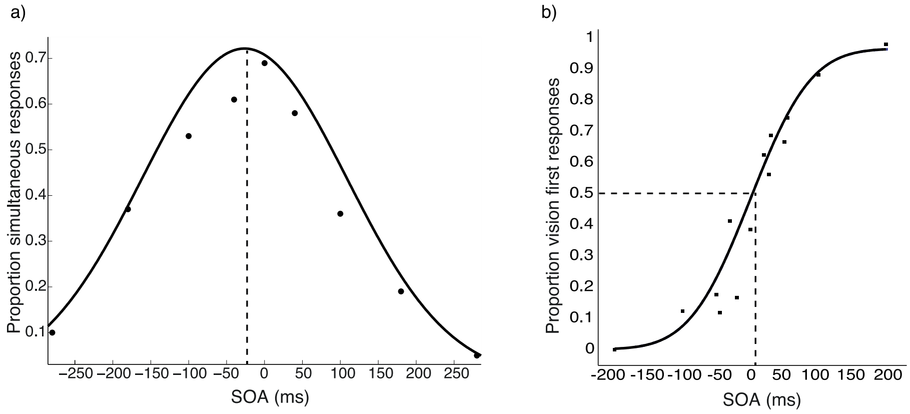


Figure 1.1. Test phase results for SJ and TOJ estimates. The differences in PSS for a) SJ for audiovisual stimuli where PSS is represented as the mean of the normal Gaussian fit and b) for TOJ where PSS is calculated as the SOA at half height of the cumulative Gaussian fit.

This effect was first demonstrated for asynchronies between a brief tone and a flashed ring of light using SJ and TOJ (Fujisaki et al., 2004; Vroomen et al., 2004). For asynchronies up to 235ms, Fujisaki et al. showed that the PSS changes proportionally to the asynchrony and the window of simultaneity becomes wider. Moreover, the range of asynchronies for which stimuli appear related also widens in the direction of the exposed asynchrony. Exposure also induces an aftereffect when stimulus and task are different during testing, like in the case of switching ear of sound presentation after exposure or judging an ambiguous stream/bounce configuration (Sekuler, Sekuler, & Lau, 1997). The recalibration of perceived synchrony demonstrates a flexibility of our perceptual system to adapt to the statistics of the environment. In this instance recalibration can be described as a temporal updating such that the conflict is removed and the sensory system containing more 'bias' (Di Luca et al., 2009) is recalibrated. In this instance of a perceptual change over time, it could be that the system changes the neural processing of each input separately thus realigning the two signals. A viable alternative is that this exposure leads to remapping (Di Luca et al., 2009). If the process is achieved through realignment then response latencies to audio and visual signals should show a recalibration effect, i.e., shifting so as to decrease the time delay between the signals. Recalibration of response latency is explored in Chapter 4 of this thesis. Moreover, if the process of recalibration is due to first realigning and then integrating the inputs then the question is does the effect transfer to different properties of the environment such as across auditory frequency and is this transfer dependent on which signal is 'recalibrated'? Chapters 5 and 6 provide some answers to this question.

1.3.1.1 Objective 2

Multisensory information, while often redundant, is not always consistent. The second objective is to explore the change in response strategy and perception when instead of a random temporal sequence the nervous system is presented with an explicit discrepant coupling between the multisensory components for a prolonged period of time. When the nervous system is exposed to conflicting audiovisual information repeatedly it recalibrates to this discrepancy (Fujisaki et al., 2004). That is, the nervous system does not suppress one component but attempts to eliminate the temporal discrepancy by realigning the two inputs in time. However, it is currently unclear how the realignment is achieved – which sensory signal (visual or auditory) shifts toward the other? The third, fourth and fifth chapters of this thesis explore whether it is possible to predict which signal should be resolve the temporal discrepancy and recalibrated – vision or audition (Chapter 4)? The prediction is that if the information is redundant it should be possible to predict the change in processing online.

However, there is no established model for this process (see Zaidal, Turner, & Angelaki, 2011) and Chapter 5 was designed to address this lack of a model. It has been suggested that recalibration should affect the perception of all stimuli in the modality showing adaptation, irrespective of whether the non-adapted stimuli would be redundant to the adapted event. The final aspect of recalibration that is explored under the second objective is to understand whether recalibrational shifts transfer to a within-modality change in stimulus? The measures used in this thesis to explore the recalibrational effect are: the response latency of the system and the perceived simultaneity of audiovisual events.

Temporal recalibration to an explicit (though perhaps imperceptible) discrepancy between redundant audiovisual signals changes our understanding of what the system is cable of doing in order to maintain a robust representation of redundant though conflicting multisensory input.

1.4 Multisensory integration for redundant cue temporal properties

Although the multisensory processes underlying spatial integration of redundant cues are quite well known (Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; van Beers et al., 1999; for a review, see Alais et al., 2010), those underlying temporal processing remain relatively poorly understood (Eagleman, Tse, Buonomano, Janssen, Nobre, & Holcombe, 2005; Ivry & Schlerf, 2008). The psychophysical data to date suggest that separate mechanisms may be responsible for estimating temporal intervals on different time scales (microseconds, seconds, hours, days etc., see Buhusi & Meck, 2005; Buonomano & Karmarker, 2002). The scale most appropriate for the majority of perceptual processes is the millisecond to second range, as it is the critical timeframe for essential processes such as motor control, speech perception, detection of motion, etc. However, as most events and objects in the environment simultaneously provide temporal information to more than one of our senses, it is important to understand the integration of temporal information across the senses.

1.4.1 Cue combination: The same for time and space?

The terms 'temporal processing' and 'time perception' encompass a broad range of phenomena, including perceived simultaneity, temporal order, and the perception of event duration. Research on audiotactile temporal order judgments (Ley, Haggard, & Yarrow, 2009) has established that for this phenomenon the integration strategy adopted by the nervous system does not differ to that used for spatial estimates. Estimating other temporal properties like event duration – returning to our example, how long did the visual and auditory articulation last – has produced inconclusive and puzzling results (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011). In order to address these differences it is important to explore the problem faced by the nervous system when making its estimate.

To be able to discriminate the time elapsed between events or integrate duration information across the senses, our nervous system needs to overcome the problem that there are no receptors dedicated to coding temporal information. To avoid making this process noisier than it need be, the nervous system could recruit/combine multiple sources of sensory input shown to optimally reduce uncertainty for other sensory estimates. If the nervous system is optimal, per se, then uncertainty about the duration of the event should decrease for multiple cues relative to either cue alone. Moreover physical conflicts in the duration between the multiple cues should be resolved according to the less noisy estimate, vision or audition. The question is whether it is possible for our nervous system to incorporate knowledge about temporal uncertainty to optimize its representation of the time elapsed between events according to reliability of environmental input?

1.4.1.1 Objective 3

The third objective of this thesis is to understand how the nervous system combines redundant cues to event duration. For spatial properties of the environment there is unequivocal evidence that redundant information is combined by a weighted averaging strategy as discussed in Section 1.1.2. Yet for temporal properties the evidence is not so clear. In order to clarify and addressed this, the final experimental chapter (Chapter 7) explores whether the nervous system applies this optimal strategy of averaging to temporal properties despite the fact that time poses unique challenges for our senses. Redundancy allows the nervous system to decrease the uncertainty in its estimate of size, position, orientation etc. However, duration estimation is a particularly challenging perceptual estimate, set apart from other sensory estimates as it is defined only once the event is finished and the sensory information is thus no longer available to the senses at the time the estimate is formed. Results from the scant previous studies would appear to indicate that perceived duration is driven by auditory input, which forces the nervous system to adopt a less flexible approach of winner-take-all. But in this instance characterised by auditory driving (e.g., Burr et al., 2009a; Welch & Warren, 1980). In the final experimental chapter of this thesis the influence of redundancy on multisensory integration challenges these previous findings.

The experimental approaches adopted in the current dissertation are primarily behavioural. However, the behavioural data are couched in terms of both physiological mechanisms and then, in later chapters, modelled according to the more computational approach. The aspect of the physiological approach that is implemented is more theoretical and addresses whether behaviour can be linked to its own unique physiological process. The behavioural data presented in the first two chapters will be discussed in light of potential physiological mechanisms. The second approach is that described in the current section whereby formal mathematical descriptions of the observed behaviour also make it possible to predict the outcome of untested circumstances. In Chapters 4, 5, 6 and 7 the experiments and analyses are guided by assumptions governing this latter approach.

Chapter 2: Multisensory interactions in space: Extending the Colavita visual dominance effect to touch

2.1 Introduction

At a sensory level, the means by which the world comes to be organized into negotiable events and objects is a far from trivial problem, since stimulus information is typically ambiguous in terms of what properties and cues belong together. This is further complicated by the fact that multiple sensory channels can process the incoming information. Moreover, it has been observed that multisensory interactions depend upon the context in which the cues are processed and perceived (e.g., Ma, Beck, & Pouget, 2008). The problem addressed in the current chapter is one that arises when the multisensory cues are complementary and observers are required to make speeded discrimination responses to them. It has been previously suggested that the cues compete for representation in awareness (see Chapter 1.2.4). Given the enormous amount of information available at any one moment in the environment, the cues available to our senses need to be selected, so that only relevant parts of the world are represented for action. Furthermore, the incoming information may be parsed in different ways according to our intended behaviour. In fact, many species (including birds and humans) demonstrate visually-dominated behaviour in spite of the fact that multiple sources of sensory information are available (Koppen & Spence, 2007a, b, c; Sinnett, Soto-Faraco, & Spence, 2008), at least under certain conditions.

Previous research investigating audiovisual interactions in humans has demonstrated that people fail to report the presence of the auditory component of suprathreshold audiovisual targets significantly more often than they fail to detect the visual component in speeded response tasks (Koppen & Spence, 2007a, Koppen & Spence, 2007b, Sinnett, Spence, & Soto-Faraco, 2008). As described in Chapter 1.1, this phenomenon is known as the Colavita effect (Colavita, 1974; see Spence, 2009, for a review). However, in looking at the possible reasons for this extinction of an auditory response, one natural question that arises is, is the phenomenon unique to audiovisual interactions? In this chapter visuo-tactile interactions using the same speeded response paradigm are investigated.

It is now well-established that our senses do not act independently of one another as once supposed (e.g., Brewster, 1832; Heymans, 1899; Jacobson, 1911; Stratton, 1899). Over the last few decades, a great deal of empirical research has documented how the different senses interact with one another (see Bremer, Lewkowicz, & Spence, 2012; Alais et al., 2010, for reviews). This research has led both to a better understanding of the conditions governing the interaction between the senses (e.g., Ernst & Bülthoff 2004; Spence, 2007), as well as to a better understanding of the physiological processes and neural substrates underlying many of these interactions (e.g., Meredith 2004; Stein et al., 2004).

One of the most interesting findings to have emerged from much of the research on multisensory perception is that vision tends to dominate over the other senses under certain laboratory conditions (e.g., when the multiple cues are complementary), if no manipulation of stimulus reliability is made. As a consequence, multisensory perception is often influenced most strongly by what we see (e.g., see Bertelson & de Gelder 2004; Rock & Harris 1967; Welch & Warren 1980, 1986, for reviews). For example, it has been shown that vision can dominate over audition in localization judgments (Bertelson & Aschersleben 1998; Pick et al., 1969) though the reverse is also possible (Alais & Burr, 2004), and over proprioception when trying to determine the position of one's limbs in space (e.g., Botvinick & Cohen, 1998; Ehrsson et al., 2004; Gallace & Spence, 2005; Hay et al. 1965; Pavani, Spence, & Driver, 2000). Vision has also been shown to be the dominant modality in a number of other animal species as well (e.g., Partan & Marler, 1999; Randich, Klein, & LoLordo, 1978; Uetake & Kudo, 1994).

As yet, however, there is no universally accepted explanation as to why, and under which circumstances, vision dominates over the other senses. One influential – but now out-dated – explanation was offered by Posner et al. (1976): They proposed that visual stimuli are not as automatically alerting as stimuli presented in other sensory modalities, such as touch and audition. Consequently, they argued that attentional resources are biased toward vision in order *"to compensate for the low alerting capability of visual signals"* (Posner et al., 1976, p. 161). Although Posner et al.'s attentional account of visual dominance has received a great deal of support in the literature over the years (see e.g., Klein, 1977), it cannot (at least, in its original form) easily explain those conditions in which other sensory modalities have been shown to dominate over vision (e.g., Guest & Spence, 2003; Lederman et al., 1986; Shams et al., 2000; though see Spence, 2009, for an updated attentional account of sensory dominance).

A prominent example of vision dominating over another sense (audition) has been provided by research on the Colavita visual dominance effect. This refers to the phenomenon whereby observers fail to respond to the auditory component of a multisensory stimulus significantly more often than they fail to respond to the visual component when they have to make speeded responses. In his original series of experiments, Colavita (1974) presented participants with visual and auditory stimuli (i.e., lights and tones) to which they had to make speeded responses using one of a pair of keys for the "light" detection response and the other key for the "tone" detection response, respectively. In the majority of trials, only one of the two stimuli (either a light or a tone) was presented. That is, most trials were unisensory. However, intermixed amongst these unisensory trials were a small number of multisensory trials in which a light and tone were presented simultaneously. Colavita deceived his participants into believing that the multisensory trials had occurred "accidentally" and did not instruct them as to how to respond on such trials. Colavita found that in 49 out of 50 multisensory trials (across the 10 participants tested), participants responded to the multisensory stimulation by pressing only the light key, and failing to respond to the auditory stimulus. Another striking observation to emerge from Colavita's original study was that the participants actually reported being unaware of the occurrence of the auditory stimulus on nearly one third of those trials. This pattern of results was even found when the subjective intensity of the tone was increased to double that of the light. Colavita argued that these findings reflected the strong dominance of vision over audition, at least under conditions where observers have to make speeded responses to simultaneously-presented pairs of auditory-visual stimuli.

Since Colavita's (1974) original studies, many other researchers have also demonstrated the Colavita visual dominance effect using a variety of different experimental set-ups (e.g., Colavita et al., 1976; Colavita & Weisberg, 1979; Shapiro & Johnson, 1989), as well as a variety of somewhat different behavioural measures (e.g., Egeth & Sager, 1977; Quinlan, 2000). Colavita and others (e.g., Colavita & Weisberg, 1979; Egeth & Sager, 1977) have tried to explain this particular visual dominance effect using Posner et al.'s (1976) attentional account (see above). It should, however, be noted that the magnitude of the visual dominance effect reported in more recent studies has tended to be much smaller than that reported by Colavita (1974) in his original research, presumably due to the elimination of many of the biases (and experimenter expectancy effects) that may have been present in the earlier research (see Spence, 2009). Somewhat surprisingly given the increasing general interest in studying crossmodal interactions, there have been no attempts so far to extend the Colavita effect to other pairs of sensory modalities than vision and audition. In a provocative paper published in Nature, Gregory (1967) claimed that touch may be a far more "primitive" sensory modality as compared to vision and audition, requiring only minimal processing capacity. In this chapter, the experiments are designed to test whether this peculiarity of the tactile modality might play an important role in multisensory processing when visual and tactile stimuli are presented together, and thus compete for access to a person's cognitive resources and/or to response system. In addition, simultaneous visual and tactile stimulation are increasingly being combined in many modern multimedia applications, for example, in communication technologies (e.g., mobile phones) or for warning signals and multisensory information displays (e.g., Ho & Spence, 2005; Van Erp & Van Veen, 2004; see Gallace et al., 2007, for a review) thus making a better understanding of how these two senses interact in terms of information processing particularly relevant at the present time.

The primary aim of the work presented in this chapter was therefore to investigate whether the visual dominance that has been observed in previous studies of the visual-auditory Colavita effect would also prevail when visual stimuli were paired with tactile stimuli (see Experiments 2.1 & 2.2). The importance of the question of whether or not vision dominates over touch is rendered even more interesting by the fact that tactile stimuli have been claimed to be of great adaptive importance for humans, given that they are thought to signal direct contact between external stimuli and the body surface (i.e., a potentially dangerous situation requiring rapid reactions; see Gregory, 1967). On the basis of this observation, one might therefore expect that vision would not dominate over tactile perception. If, however, an 'extinction' of tactile inputs were to be observed in the presence of simultaneous visual stimulation, then this may help to shed some light upon (and perhaps provide an efficient means of exploring) crossmodal extinction-like phenomena in neurologically-normal observers (Bender et al., 1952; Costantini et al., 2007). Investigating the Colavita effect for various combinations of sensory modalities might therefore provide some important insights into the putative mechanisms underlying the conscious access and response to information when a competition for processing resources (and/or neural representations) occurs between different sensory modalities (see Costantini et al., 2007; Gallace & Spence, 2007).

A second major goal of the experiments reported in this chapter was to test the extent to which the Colavita visual dominance effect is dependent upon the relative spatial positions from which the visual and tactile stimuli are presented (Experiment 2.3). Indeed, previous research into the influence of spatial coincidence on crossmodal interactions has demonstrated that the interactions between stimuli in different sensory modalities are sometimes, but importantly not always, weaker when the stimuli are presented from different spatial positions as compared to when they are presented from the same position (see Congedo et al., 2006; Gephstein, Burge, Ernst, & Banks, 2005; see Maravita et al., 2003; Spence & Driver, 2004; Spence, 2007, for reviews). Finally, in Experiment 2.4, the contribution of stimulus intensity to the visual dominance ove touch was checked by perceptually matching the intensities of the stimuli in the two modalities.

2.2 Experiment 2.1: Transfer to vision and touch

Previous research on visuotactile interactions has shown that while vision dominates over touch under certain conditions (e.g., Gibson, 1943; Rock & Harris, 1967; Rock & Victor, 1964), touch dominates over vision in many others (e.g., Spence, 2009, for review). Experiment 2.1 was therefore designed to investigate whether vision would dominate over touch in a speeded discrimination task setting using the Colavita paradigm. In particular, the aim was to determine whether observers would tend to respond to the visual element of multisensory stimuli (visual-only responses) significantly more frequently than they responded to the tactile element (tactile-only responses). To look at the difference across sensory pairings for this effect the experimental design was directly based on that reported by Koppen and Spence (2007a, b, c) in their recent studies of the 'more traditional' audiovisual Colavita effect.

2.3 Methods: Experiment 2.1

2.3.1 Observers

Ten naïve observers (8 male, 2 female) aged between 19 and 29 years (mean age of 23 years) took part in Experiment 2.1, which lasted for approximately 30 minutes. All of the observers were right-handed by self-report and had normal or corrected-to-normal vision as well as normal tactile sensitivity. The observers in all of the experiments reported here received 5 pound (U.K. Sterling) gift voucher in return for taking part in each experiment. The experiments were conducted in accordance with the ethical guidelines laid down in the Declaration of Helsinki (World Medical Association, 2000).

2.3.2 Apparatus and materials

The experiments were conducted in a dimly-illuminated testing booth. The observers sat on a chair with their right index finger placed in a custom-built stimulus box, positioned centrally on a table approximately 65 cm in front of them (see Figure 2.1). The stimulus box was used to present the visual and tactile stimuli. The visual stimuli consisted of the 50 ms illumination of a yellow light emitting diode (LED), CIE (0.4, 0.47), with a luminance of 0.98 cd/m^2, positioned so as to illuminate the tip of the participant's index finger (cf. Hari & Jousmäki, 1996). Tactile stimuli were presented by means of a bone conduction vibrator (Oticon Limited, B/C 2-PIN, 100 Ohm, Hamilton, UK) with a 1.6 x 2.4 cm vibrating surface. The vibrotactile stimuli consisted of 50 ms 250 Hz supra-threshold pulses generated by a signal generator. The vibrator was mounted in a foam block placed directly below the LED. The tip of the observers' index finger was positioned directly above the vibrator, thus ensuring that the visual and tactile stimuli were presented from the same spatial position.

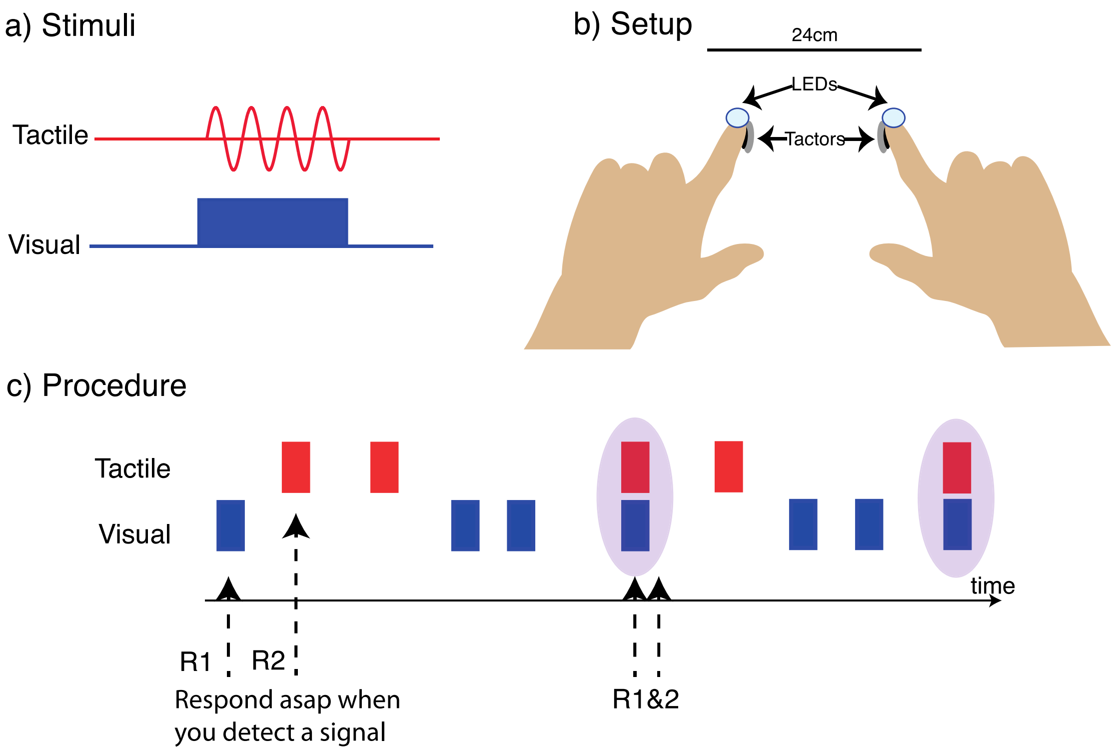


Figure 2.1. **Schematic representation of a) the signals, b) the experimental set-up and c) the procedure used in all of the experiments presented in this chapter. R1 indicates response to stimulus 1 (visual or tactile) and R2 represents response to the second stimulus.**

Responses were collected using two foot-pedals placed on the floor below the table (one placed 10 cm to the left and the other 10 cm to the right of the participant's body midline), on which the observers rested their feet. The observers were instructed to lift their toes off of one foot pedal in response to the visual targets and to lift their toes off the other foot pedal in response to the tactile targets. The allocation of the stimuli to the foot pedals was counterbalanced across observers. The observers were explicitly instructed to respond by lifting their feet off both foot pedals on the multisensory target trials. However, no specific instructions were provided to the observers as to whether or not they should lift the two foot-pedals simultaneously. White noise was presented to observers via headphones to mask any background noise.

2.3.3 Design

Each participant completed 6 blocks of 100 trials. Each block of trials consisted of 40 unisensory visual, 40 unisensory tactile, and 20 multisensory stimuli. The probability of the different stimulus types presented was the same as those reported in a number of previous studies (e.g., Egeth & Sager, 1977; Koppen & Spence, 2007a, b) and was also similar to those used by Colavita in his original research (Colavita, 1974; Colavita et al., 1976). The order of stimulus presentation was randomised within each block of trials. A block of 30 practice trials was presented before the main experimental session. The practice trials were identical to the experimental trials but were not analysed.

2.3.4 Procedure

The observers were instructed to fixate on the tip of their right index finger throughout each block of trials. The observers were randomly presented with a unisensory visual, a unisensory tactile, or a multisensory stimulus on each trial. The stimuli were presented for 50 ms and followed by a 1500 ms response interval, after which the next trial began automatically. The observers were instructed to respond as rapidly and accurately as possible. No feedback regarding the correctness of a participant's responses was provided. Note that the observers were able to make a second response, should they have wanted to, after having made their initial response (provided that they did so within the response interval before the onset of the target on the next trial).

2.4 Results: Experiment 2.1

The observers failed to make any response on 1.7% of the trials, and these trials were not included in any of the data analyses. The results are summarised in Table 2.1. The percentages of multisensory trials in which the observers only responded to one of the two stimuli (i.e., they either made a visual-only or a tactile-only response) were analysed first. The Colavita visual dominance effect would consist of observers making significantly more visual-only than tactile-only responses on the multisensory trials (see Colavita, 1974, for a similar definition, and Spence, 2009, for a review).

Table 2.1. Mean error rates for the unisensory tactile, unisensory visual, and multisensory target stimuli presented in Experiments 2.1 and 2.2. Observers could make one of two types of error (tactile-only or visual-only response) on the multisensory trials. Mean RTs (ms) for correct responses to unisensory tactile, unisensory visual, and multisensory target stimuli are presented. Multisensory tactile and visual responses refer to the RTs to the tactile and visual components of the multisensory target stimuli (i.e., tactile or visual key presses). Standard errors are shown in parentheses.

	Experiment 2.1	Experiment 2.2
Stimulus probability	40T; 40V; 20VT	35T; 35V; 35VT
Error rates (%)		
Unisensory tactile	8.1 (1.8)	6.3 (1.4)
Unisensory visual	6.3 (1.6)	6.0 (1.5)
Multisensory cues		
Tactile-only	5.2 (1.4)	3.1 (1.5)
Visual-only	15.6 (3.5)	6.5 (1.4)
Reaction times (ms)		
Unisensory tactile	572 (28)	547 (37)
Unisensory visual	525 (40)	512 (35)
Multisensory cues		
Tactile-only	634 (33)	621 (45)
Visual-only	618 (50)	613 (44)

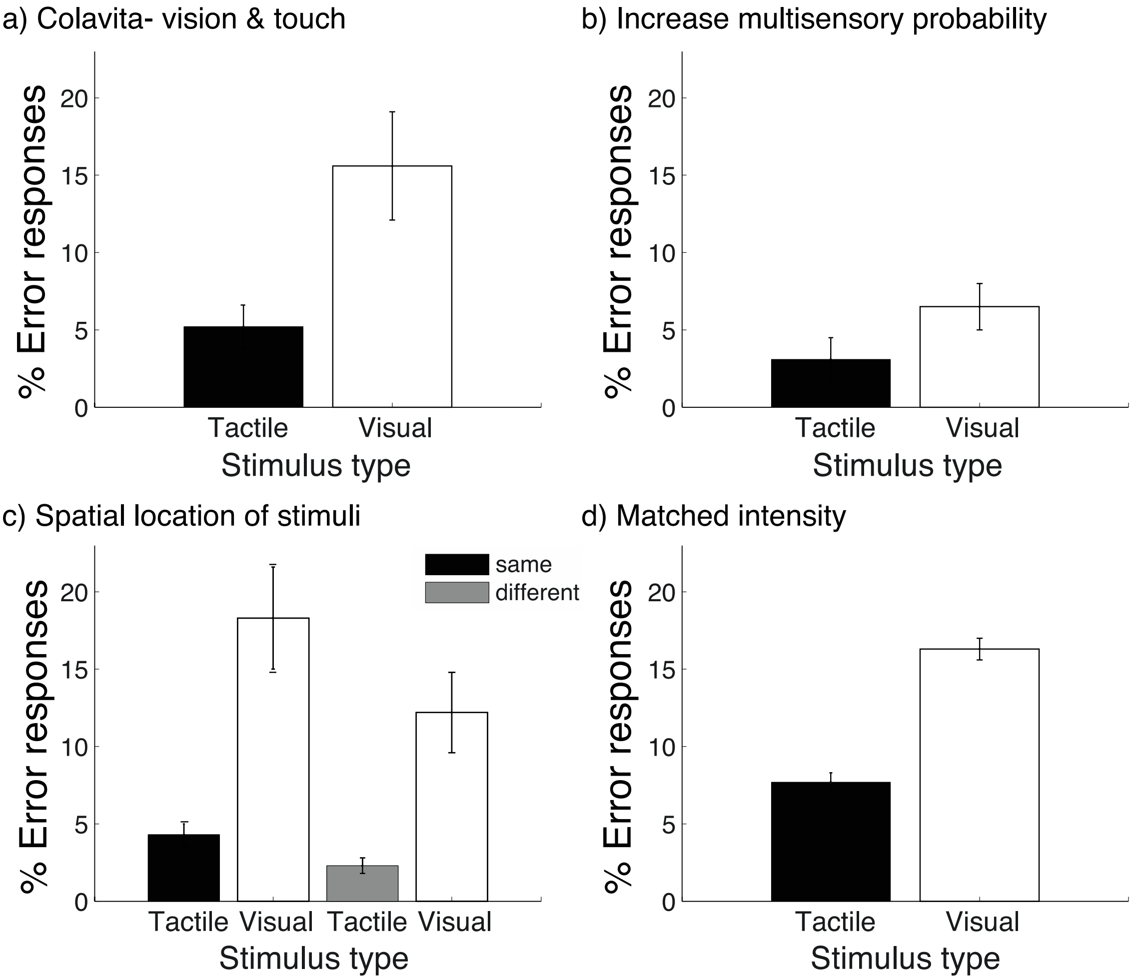


Figure 2.2. Percentage of visual-only and tactile-only responses on multisensory trials in Experiments 2.1 (a), 2.2 (b), 2.3 (c), and 2.4 (d). The error bars represent the standard errors of the means.

2.4.1 Error data

The percentages of multisensory trials in which the observers only responded to one of the two stimuli were compared using a paired-samples t-test (note that all of the t-tests reported in this chapter are two-tailed). This comparison revealed that the observers made significantly more visual-only (M=15.6%) than tactile-only responses overall (M 5.2%; t(9)=2.61, p=0.028; see Figure 2.2a).

A one-way repeated measures (RM) ANOVA was conducted on the error data with the factor of Trial type (unisensory visual, unisensory tactile or multisensory). In this and all subsequent ANOVAs, Greenhouse-Geisser corrections were made whenever the assumption of sphericity was violated. The analysis of the error data revealed a significant main effect of Trial type [F(1,9)=11.56, p=0.005], which was further examined by means of pairwise comparisons (Bonferroni-adjusted for multiple comparisons). The results showed that the observers made significantly more errors on the multisensory trials (M=20.8%) than on either the unisensory visual trials (M=6.3%, p=0.019) or on the unisensory tactile trials (M=8.1%, p=0.022). There was no significant difference between the latter two conditions (p=0.75).

2.4.2 RT data

The mean RTs for those trials in which the observers responded correctly were analyzed using an ANOVA with the factor of Stimulus modality (visual vs. tactile) and Trial type (unisensory vs. multisensory). This analysis revealed a main effect of Trial type [F(1,9)=99.58, p<0.001], with observers responding significantly more slowly on the multisensory trials (M=626 ms) than on the unisensory trials (M=549 ms), as expected. The main effect of Stimulus modality was, however, not significant, nor was the interaction between Stimulus modality and Trial type [both Fs<1, n.s.].

2.5 Discussion: Experiment 2.1

The key finding to emerge from Experiment 2.1 was that observers made significantly more visual-only than tactile-only responses on the multisensory target trials. This result provided the first empirical demonstration of the existence of a Colavita effect between vision and touch (see Hartcher-O'Brien et al., 2008; though see Hecht & Reiner, 2009). This finding would appear to conflict with previous claims that given the particular biological relevance of tactile stimuli to humans, they should be especially capable of eliciting an immediate response (see Gregory, 1967). The magnitude of the visuotactile Colavita effect reported here (15.6% visual-only and 5.2% tactile-only responses, a 10.4% Colavita effect) was similar to that reported by Koppen and Spence (2007a) in their study of the visual dominance over audition (12.1% visual-only and 4.4% auditory-only responses, producing a 7.7% Colavita effect). Further comparison of the results of these two studies using an ANOVA with the within-participants factor of Response (visual-only vs. tactile-only) and the between-participant factor of Modality pairing (visual-auditory vs. visual-tactile) revealed a significant main effect of Response [F(1,22)=13.97, p=0.001], with the observers making more visual-only (M=13.8%) than auditory/tactile-only responses (M=4.7%), thereby showing a significant Colavita effect. Importantly, however, there was no main effect of Modality pairing, nor any interaction between these two factors [both F<1, n.s.]. These results therefore suggest that the magnitude of the Colavita effect in the visual-tactile domain was similar to that observed previously in the visual-auditory domain[1].

---

[1] A more detailed comparison of the two studies is made difficult by the fact that no specific attempt was made to match the intensity of the stimuli, either within or between these studies. It should, however, be noted that Colavita and Weisberg (1979) explicitly looked for any effect of changes in the intensity of the auditory stimulus on the magnitude of the visual dominance over audition. They found absolutely no effect of intensity changes on the magnitude of the Colavita visual dominance effect, even when the subjective intensity of the sound was doubled from one experiment to the next.

The reaction times observed in the current experiment were large, relative to choice RTs (in the range of 200-300ms). This could be indicative of the responses being the result of a decision-level process (see Hanson et al., 2009). Furthermore, the increased error rates and slower RTs observed for correct responses on the multisensory trials as compared to the unisensory trials would seem to indicate that the observers in Experiment 2.1 found it more difficult to respond to the multisensory targets than to either the unisensory visual or unisensory tactile targets. This result presumably reflects the fact that more computational/cognitive resources are required to perceive and/or respond to two simultaneously-presented stimuli than to perceive/respond to just a single stimulus, especially when the two stimuli are presented at about the same time (e.g., see Pashler, 1994), when the two stimuli are complementary and not integrated. However, there is also an alternative explanation for the increased RTs and error rates observed on the multisensory trials in Experiment 2.1. In particular, it has been shown previously that observers in speeded discrimination tasks often respond more rapidly and more accurately to a target when it has been preceded by another stimulus of the same kind than when it has been preceded by a different stimulus – an effect known as "repetition priming" (e.g., Bertelson, 1961; Pashler & Baylis, 1991; Spence, Nicholls, & Driver, 2001; Turatto et al., 2004). As described earlier, the unisensory visual, unisensory tactile, and multisensory trials in each block of trials in Experiment 2.1 were presented in the ratio of 40:40:20 (in line with the majority of previous research in this area; e.g., see Egeth & Sager, 1977; Koppen & Spence, 2007a). However, it should be noted that due to the somewhat lower frequency with which the multisensory stimuli were presented, the probability of two multisensory trials occurring on consecutive trials was much lower than the probability with which either two unisensory visual or two unisensory tactile trials would occur consecutively. As a consequence, observers' responses to the multisensory stimuli were less likely to benefit from any repetition priming[2] effects that may have been present in the design of Experiment 2.1[3].

2.6 Experiment 2.2: Stimulus probability

Experiment 2.2 was therefore designed to test whether or not the Colavita effect observed in Experiment 2.1 was simply attributable to the relative probability with which the three types of trial were presented in Experiment 2.1. The design of this new experiment was identical to that of Experiment 2.1 with the sole exception that the relative probability of the three types of target (unisensory visual, unisensory tactile, and multisensory) was now equalized. Therefore if the effect arises from the predominantly unisensory context in which the nervous system found itself, then equalizing the unisensory and multisensory trials should diminish the effect.

2.7 Methods: Experiment 2.2

2.7.1 Observers

Ten naïve observers (6 male, 4 female) aged between 20 and 30 years (mean age of 23 years) took part in Experiment 2.2, which lasted for approximately 30 minutes.

---

[2] Note that this argument makes the assumption that the two elements of a multisensory trial are perceived as a pair of stimuli that are associated with their own unique response, i.e., lifting both foot pedals (cf. Fagot & Pashler, 1992; Koppen & Spence, 2007a; Schumacher et al., 2001).

[3] While both stimuli were presented at a supra-threshold level, no attempt was made in this experiment to match the intensities of the visual and vibrotactile stimuli (cf. Koppen & Spence, 2007a, b, c; Sinnett et al., 2007; Spence et al., 2001b). Thus, it is possible that the visual and tactile stimuli may have been presented at different subjective intensities.

2.7.2 Apparatus, materials, design, and procedure

These were exactly the same as in the previous experiment with the sole exception that an equal number of unisensory visual, unisensory tactile, and multisensory trials were now presented in each block of trials. There were 30 practice trials (10 unisensory visual, 10 unisensory tactile, and 10 multisensory), followed by 6 blocks of 105 trials (35 unisensory visual, 35 unisensory tactile, and 35 multisensory).

2.8 Results: Experiment 2.2

Trials in which the observers failed to respond (less than 1% of the trials overall) were not included in the subsequent data analyses. The results of Experiment 2.2 are highlighted in Table 2.1.

2.8.1 Error data

A paired samples t-test comparing the percentages of visual-only and tactile-only responses on the multisensory trials revealed a significant difference [t(9)=2.28, p=0.049], with the observers once again making significantly more vision-only responses (M=6.5%) than touch-only responses (M=3.1%), as shown in Figure 2.2b. An analysis of the error data with a one-way repeated measures ANOVA with Trial type (unisensory visual vs. unisensory tactile vs. multisensory) as the within-participant factor revealed no significant main effect of Trial type [F(2,18)=1.55, p=0.250].

2.8.2 RT data

Mean RTs for correct responses were compared using a 2 × 2 repeated measures ANOVA with the within-participants factors of Stimulus modality (visual vs. tactile) and Trial type (unisensory vs. multisensory). This analysis revealed a significant main effect of Trial type [F(1,9)=43.67, p<0.001], with the observers responding more slowly on the multisensory trials (M=617 ms) than on the unisensory trials (M=530 ms), just as had been reported in the previous experiment. Although the main effect of Stimulus modality was not significant [F<1, n.s.], the main effect of Trial type was qualified by a significant interaction with Stimulus modality [F(1,9)=7.20, p=0.025]. Analysis of the simple main effects showed that Trial type had a significant effect on people's responses to both visual stimuli [M=613 ms and 512 ms for multisensory and unisensory trials, respectively; F(1,9)=70.00, p<0.001] and tactile stimuli [M=621 ms and 547 ms for multisensory and unisensory trials respectively; F(1,9)=21.47, p=0.001]. The difference between the two trial types was, however, significantly smaller for the tactile stimuli (MD=74 ms) than for the visual stimuli (MD=101 ms), t(9)=-2.68, p=0.025.

2.9 Discussion: Experiment 2.2

The results of Experiment 2.2 once again revealed a significant Colavita visual dominance effect over touch. That is, the observers made significantly more vision-only responses than touch-only responses on the multisensory trials. This result demonstrates that simply matching the stimulus probabilities or trial types cannot eliminate the Colavita effect. However, it is important to note that visual inspection of the results of Experiments 2.1 and 2.2 (compare Figures 2.2a & 2.2b) suggests that the effect was somewhat smaller in this experiment than in Experiment 2.1. To further examine how the relative probability of the three different trial types affected the Colavita effect, the responses given on the multisensory trials in Experiments 2.1 and 2.2 were compared directly. An ANOVA was conducted on the error data from the multisensory trials with the within-participant factor of Response (visual-only vs.

tactile-only) and the between-participants factor of Experiment (2.1 vs. 2.2). The analysis revealed a significant main effect of Response [F(1,18)=10.51, p=0.005], demonstrating that the observers made significantly more vision-only (M=11.1%) than touch-only responses (M=4.2%), as expected, thus demonstrating a robust Colavita visual dominance effect. There was also a significant main effect of Experiment [F(1,18)=6.53, p=0.020], with observers making more errors overall in the multisensory trials of Experiment 2.2 (M=10.4%) than in the equivalent trials of Experiment 2.1 (M=4.8%). Crucially, however, the analysis did not reveal any interaction between these two factors [F(1,18)=2.73, p=0.16], thus showing that there was no significant difference in the magnitude of the Colavita visual dominance effect in the two experiments reported so far. Hence, the results of Experiment 2.2 show that the Colavita visual dominance effect reported in Experiment 2.1 was not simply an artefact of the lower probability of multisensory targets, since the magnitude of the effect was unaffected by equalizing the probability with which each of the three stimulus types was presented. This is not, of course, to say that the Colavita effect might not be significantly decreased, or even completely eliminated by increasing the probability of multisensory stimuli still further, so that multisensory stimuli were much more frequent than either the unisensory visual or unisensory tactile stimuli (see Koppen & Spence, 2007c, on this point).

2.10 Experiment 2.3

The two experiments presented so far in this thesis provide the first empirical demonstration that vision can dominate over touch in the Colavita task (see Hartcher-O'Brien et al., 2008). In the following experiment, the idea was to investigate whether the introduction of a spatial separation between the targets on the multisensory trials would have a significant effect on the magnitude of the Colavita effect. In Experiments 2.1 and 2.2, the visual and tactile stimuli were always presented from exactly the same central position (i.e., at fixation). In Experiment 2.3, the stimuli were now presented from one of two possible positions, one placed to either side of central fixation (with an equal number of stimuli being presented from either position). Given that in Experiments 2.1 and 2.2 all of the stimuli had been presented at fixation (i.e., at the focus of a participant's overt attention), it may be possible that presenting the stimuli from a more peripheral location (i.e., 12.5° from central fixation) might, if anything, reduce the dominance of vision over touch and hence perhaps reduce the magnitude of the Colavita visual dominance effect.

2.11 Methods: Experiment 2.3

2.11.1 Observers

Sixteen naïve observers (10 male, 6 female) aged between 18 and 45 (mean age of 25 years) took part in Experiment 2.3, which lasted for approximately 40 minutes. 14 were right-handed and 2 left-handed by self-report.

2.11.2 Stimuli

These were the same as in the two previous experiments with the following exceptions: Two identical stimulus boxes, each containing an LED and a vibrotactile stimulator were now used to present the stimuli. The two stimulus boxes were placed approximately 65 cm in front of the participant. They were located equidistant (12.5°) to the right and left of central fixation. The observers placed their right index finger in the right stimulus box and their left index finger in the left box. The two locations from which the targets could be presented (i.e., from the fingertips of observers' right and left index finger) were placed at the same elevation as the fixation point).

2.11.3 Design and procedure

Each participant completed 8 blocks of 100 trials. Each block consisted of 40 unisensory visual, 40 unisensory tactile, and 20 multisensory trials. Note that given that varying the frequency with which the multisensory stimuli were presented had been shown in previous experiments to have no significant effect on the magnitude of the Colavita effect, the trial frequencies used in the first experiment were again used here, which resulted in a numerically somewhat larger number of multisensory errors for analysis. The unisensory targets were presented equally often from either side of fixation. In half of the multisensory trials, the two stimuli were presented from the same spatial position, while in the other half of the trials they were presented from different positions (one to either side of fixation). The four possible combinations of stimuli (i.e., visual left – tactile left, visual right – tactile right, visual left – tactile right, and visual right – tactile left) were presented equally often in each block of trials. The various stimuli were presented in a randomised order for each participant. Prior to the main experiment observers received 40 practice trials. These practice trials consisted of 16 unisensory visual, 16 unisensory tactile, and 8 multisensory trials. The procedure was identical to that adopted in the two previous experiments. The observers in this experiment were also asked to ignore the position of the target stimuli and to respond only to the sensory modality in which they were presented (i.e., whether they saw a light, felt a vibration, or experienced both stimuli). They were also instructed to keep their eyes on the central fixation point throughout the experiment.

Table 2.2. Mean error rates for unisensory tactile, unisensory visual, and multisensory target stimuli, with the two types of multisensory errors (tactile-only and visual-only responses) presented below the average multisensory error rates in Experiment 2.3. Mean RTs (ms) for correct responses to unisensory tactile, unisensory visual, and multisensory target stimuli, with the RTs to the tactile and visual components of the multisensory target stimuli presented directly below the average multisensory RTs. Standard errors are shown in parentheses.

	Unisensory cue	Multisensory cues	
Response		Spatial location	
Error rates (%)		Same	Different
Unisensory tactile	9.3 (1.6)		
Unisensory visual	8.3 (1.1)		
Multisensory cues		**22.6 (3.4)**	**14.5 (2.9)**
Tactile-only		4.3 (0.7)	2.3 (0.5)
Visual-only		18.3 (3.3)	12.2 (2.6)
Reaction times (ms)			
Unisensory tactile	697 (23)	547 (37)	
Unisensory visual	629 (22)	512 (35)	
Multisensory cues			
Tactile-only		644 (32)	673 (30)
Visual-only		666 (20)	670 (20)

2.12 Results: Experiment 2.3

Observers failed to respond on 1.8% of trials overall, and these trials were not included in the subsequent data analyses. The results of Experiment 2.3 are highlighted in Table 2.2.

2.12.1 Error data

The percentage of multisensory trials in which the observers responded only to the visual stimulus or only to the tactile stimulus, were compared for the two different trial types (same vs. different position). An ANOVA was conducted with the factors of Response (visual-only vs. tactile-only) and Spatial position (same vs. different). This analysis revealed a main effect of Response [F(1,15)=19.66, p<0.001], with observers making significantly more visual-only responses (M=15.2%) than tactile-only responses (M=3.3%), just as reported in the two previous experiments. The analysis also revealed a main effect of Spatial position [F(1,15)=12.72, p=0.003], with observers being significantly more likely to erroneously make only a single response when the two stimuli were presented from the same spatial position (M=11.3%) than when they were presented from different positions (M=7.2%). The ANOVA also revealed a significant interaction between response and spatial position [F(1,15)=4.16, p=0.048]. An analysis of the simple main effects revealed that observers made more visual-only than tactile-only responses on the multisensory trials both when the two stimuli were presented from the same position [F(1,15)=18.41, p=0.001], and also when they were presented from different positions [F(1,15)=17.06, p=0.001]. This difference was, however, significantly larger for those multisensory trials in which the stimuli were presented from the same position (MD=14.0%) than for those multisensory trials in which the two stimuli were presented from different positions (MD=9.8%; t(15)=2.15, p =0.048; see Figure 2.2c).

The overall error rates were analysed using a one-way RM ANOVA with the factor of Trial type (unisensory visual, unisensory tactile, multisensory same position, multisensory different positions). This analysis revealed a significant main effect [F(3,45)=9.73, p=0.002]. Subsequent pairwise comparisons (Bonferroni-corrected for multiple comparisons) showed that the observers made significantly more errors on the multisensory trials when the stimuli were presented from the same position (M=22.6%) than on those multisensory trials where the stimuli were presented from different positions (M=14.5%, p=0.018), or on either unisensory tactile (M=9.3%, p=0.015), or unisensory visual trials (M=8.3%, p=0.009). None of the differences between the other three trial types were significant.

2.12.2 RT data

The mean RT data from those trials in which the observers responded correctly were examined using a RM ANOVA with the within-participant factors of Stimulus modality (visual vs. tactile) and Trial type (unisensory vs. multisensory same position vs. multisensory different positions). This analysis failed to reveal a significant main effect of either factor [both F<1, n.s.]. The interaction between Stimulus modality and Trial type was, however, significant [F(2,30)=9.43, p=0.001]. A subsequent analysis of the simple main effects revealed that the observers in Experiment 2.3 responded more slowly to tactile than to visual stimuli on the unisensory trials [M=697 vs. 629 ms, respectively; F(1,15)=5.05, p=0.04], but not on the multisensory trials (neither when the stimuli were presented from the same spatial position, nor when they were presented from different positions [both F(1,15)<1, n.s.]).

2.13 Discussion: Experiment 2.3

The results of Experiment 2.3 demonstrate the existence of a significant Colavita visual dominance effect no matter whether the visual and tactile stimuli were presented from the same or from different spatial positions (separated by 25° of visual angle). However, the magnitude of the Colavita visual dominance over touch was found to be significantly larger when the two stimuli were presented from the same location than when they were presented from different locations (cf. Congedo et al., 2006; Gephstein et al., 2005, for similar results from studies using very different experimental paradigms). The spatial modulation of the visual-tactile Colavita effect therefore appears to bear some similarity to the spatial modulation of the visual-auditory Colavita effect reported recently by Koppen and Spence (2007b; see also Shapiro & Johnson, 1987).

2.14 Experiment 2.4: Matching stimuli for intensity

It seemed possible that the visual dominance effect observed in Experiments 2.1 through 2.3, under conditions of multisensory visuotactile stimulus presentation, might have resulted from a simple difference in the perceived intensity of the unisensory stimuli that were presented, rather than necessarily representing a more fundamental difference in information-processing between the senses. That is, it was possible that the intensity of the visual stimulus might (serendipitously) have been perceived by observers as having been much greater than that of the tactile stimulus. As a consequence, the 'extinction' of the tactile stimulus by vision on a proportion of the trials in the three experiments reported so far may have been solely attributable to a mismatch in the relative intensity of the stimuli used. Therefore, in order to rule out the intensity-based explanation of the visual dominance over touch reported here, a fourth experiment was conducted in which the observers had to explicitly match the perceived intensity of the visual and tactile stimuli at the start of their experimental sessions. They were asked to match the intensity of the tactile signal to that of a fixed-intensity LED presented at threshold. Note that matching the stimulus intensities of the signals presented in different modalities is a procedure that is not often undertaken due to uncertainty concerning the question of how exactly stimuli can (or should) be matched crossmodally (see Spence et al., 2001, on this point). However, it is perhaps worth noting none of the observers reported any conceptual difficulty with performing this task.

2.15 Methods: Experiment 2.4

2.15.1 Observers

Thirteen naïve observers (5 male, 8 female) aged between 18 and 45 (mean age of 25 years) took part in Experiment 2.4. The experiment lasted approximately 50 minutes.

2.15.2 Apparatus and materials

These were the same as in Experiment 2.1 with the following exceptions: The visual stimulus now consisted of a variable-intensity white LED. The intensity of the tactile stimulus was adjusted manually by the participant at the start of the experiment by means of an analogue amplifier. Each participant's detection threshold for the visual stimulus was measured at the beginning of the experimental session. Detection responses were made using the keyboard, after which observers completed the manual matching procedure followed by the Colavita task using the foot pedal responses as in Experiments 2.1 – 2.3).

2.15.3 Visual threshold procedure

The observers were dark-adapted for five minutes at the beginning of their experimental session. The luminance of the LED ranged between 0.64-0.59cd/m^2 using the double-random staircase procedure (three-up/one-down; see Cornsweet, 1962; Koppen, Levitan, & Spence, 2009; step size of approximately 0.003 cd/m^2). Each staircase was terminated upon the 12th reversal of the sequence with the 75% detection threshold calculated from the average of the preceding eight reversals. The stimulus intensities at which observers were able to detect the visual stimulus on 75% of the trials were measured in separate blocks of experimental trials. On each trial, the target was followed by a 1550ms response interval (i.e., successive stimuli were always separated by 1600ms). The observers were informed that the intensity of the visual stimuli would vary from trial to trial, and that they should press the press the 'm' key whenever they detected a target, and make no response if no target was detected. On average, it took 82 trials (S.E.=7) to establish the visual threshold. The average visual threshold was 0.60cd/m^2 (S.E.=0.08 cd/m^2).

2.15.4 Intensity matching procedure

The visual stimulus (whose intensity had been set to each participant's individual detection threshold) was held constant throughout the matching procedure. Both the visual and tactile stimuli were presented simultaneously for 50 ms with the tactile stimulus presented at the same initial intensity for all observers. On the first trial they were told to use the volume knob on the amplifier in order to sample the entire range of vibrotactile intensities. They were instructed that the visual stimulus intensity would not change and that they should match the intensity of the tactile stimulus until both were experienced as being equal perceptually. The resulting visual threshold and the subjectively-matched tactile intensity values were then used in the Colavita task itself.

2.15.5 Colavita procedure

Each participant completed 8 blocks of 100 trials. Each block consisted of 40 unisensory visual, 40 unisensory tactile, and 20 multisensory trials. The procedure was identical to that adopted in the three previous experiments.

2.16 Results: Experiment 2.4

The observers in Experiment 2.4 failed to make any response on 22% of the trials overall, and these trials were not included in the subsequent data analyses. The results are highlighted in Table 2.3. The percentage of multisensory trials in which the observers only responded to one of the two stimuli (i.e., they either made a visual-only or a tactile-only response) were analysed first.

Table 2.3. **Mean error rates for Experiment 2.4 for the unisensory tactile, unisensory visual, and multisensory target stimuli, with the two types of multisensory errors (tactile-only and visual-only responses) are presented below the average multisensory error rates. Mean RTs (ms) for correct responses to unisensory tactile, unisensory visual, and multisensory target stimuli, with the RTs to the tactile and visual components of the multisensory target stimuli presented directly below the average multisensory RTs. Standard errors are shown in parentheses.**

Experiment 2.4	
Matched intensity	
Error rates (%)	
Unisensory tactile	11 (5.1)
Unisensory visual	29 (7.6)
Multisensory cues	**33.5 (3.2)**
Tactile-only	16.3 (0.7)
Visual-only	7.7 (0.6)
Reaction times (ms)	
Unisensory tactile	695.6 (21)
Unisensory visual	675.1 (23)
Multisensory cues	
Tactile-only	789.5 (23)
Visual-only	712.6 (20)

2.16.1 Error data

A paired samples t-test comparing the percentages of visual-only and tactile-only responses for the multisensory trials revealed a significant difference [t(12)=3.35, p=0.005] between the percentage of visual- and tactile-only responses, with observers once again making significantly more vision-only (M=29.3%) than touch-only responses (M=11.0%), as shown in Figure 2.2d. An one-way RM ANOVA on the error data with Trial type (unisensory visual vs. unisensory tactile vs. multisensory) as the within-participants factor revealed a significant main effect [F(2,12)=11.45, p=0.003]. Subsequent pairwise comparisons revealed that the observers made significantly more errors on the multisensory trials (M=33.9%, p<0.05) than in either the unisensory-visual (M=7.7%) or unisensory-tactile trials (M=16.3%, p<0.05). There was no significant difference between the unisensory conditions (p>0.05).

2.16.2 RT data

The mean RTs for those trials in which the observers responded correctly were analysed using a RM ANOVA with the factors of Stimulus modality (visual vs. tactile) and Trial Type (unisensory vs. multisensory). This analysis revealed no main effect of Stimulus modality [F<1, n.s.]. There was, however a significant main effect of Trial Type [F(1, 12)=18.47, p<0.001], with the observers responding more rapidly on unisensory (M=660 ms) than on multisensory (M=790 ms, p<0.001) trials overall. The lack of any RT difference between the visual and tactile response latencies on those multisensory trials in which observers responded correctly, suggests that observers were not making sequential responses, i.e. responding to the multisensory target as if it were two separate unisensory stimuli.

2.17 Discussion: Experiment 2.4

The results of Experiment 2.4 demonstrate that the Colavita visual dominance effect over touch cannot be accounted for simply by any differences in stimulus intensity that might have been present in the first 3 experiments. In fact, the results of this final experiment replicate the findings of Koppen, Levitan, and Spence's (2009) study in which they also showed that the audiovisual Colavita effect still occurred when the intensity of the auditory and visual stimuli were matched. However, the results of the present study also extend previous research on the Colavita effect by suggesting that the result is much stronger when the signals are presented at, or near, threshold. Comparing the results of Experiment 2.4 to those of Experiments 2.1 – 2.3, it would appear that the magnitude of the visual dominance effect increases as the strength of the stimuli decrease and/or when the intensities of the stimuli in the two modalities were subjectively matched. Under these conditions, the Colavita visual dominance effect was observed on approximately 20% of the trials, as compared to a mean Colavita effect of 8.9% when the data from Experiments 2.1 – 2.3 is averaged.

2.18 General Discussion

The primary aim of the four experiments reported in the present chapter was to determine whether a Colavita visual dominance effect could be demonstrated when a visual stimulus was presented at the same time as a tactile stimulus. The results of all four experiments provide empirical support for the view that the Colavita visual dominance effect does indeed occur when visual and tactile stimuli are presented simultaneously. What is more, the results of Experiment 2.2 showed that this effect does not simply reflect an artefact of the relative probability of presentation of the unisensory visual, unisensory tactile, and multisensory trials, since equalizing the probability with which the three target types occurred did not have a significant effect on the magnitude of the Colavita effect that was reported. The results of Experiment 2.3 demonstrated that the visual dominance over touch was significantly larger when the stimuli were presented from the same spatial position than when they were presented from different positions (cf. Congedo et al., 2006; Costantini et al., 2007; Gephstein et al., 2005). Thus, the visual dominance over touch observed here in the Colavita paradigm would appear to reflect a robust effect, very similar to the visual dominance over audition that has been reported in a number of recent studies of the audiovisual Colavita effect (see Koppen & Spence, 2007a, b, c). Finally, the results of Experiment 2.4 revealed that the Colavita visual dominance effect over touch still occurs (in fact, it gets somewhat larger) when the intensity of the visual and tactile stimuli subjectively matched on a participant-by-participant basis at the start of each participant's experimental session.

According to the attentional interpretation of sensory dominance, vision dominates because of a strategic bias by people (at least in psychology experiments) to attend more to vision in order to compensate for its poor alerting capacity (Posner et al., 1976). Given that it has been argued that tactile stimuli, just like auditory stimuli, are intrinsically more alerting than visual stimuli (see Posner, 1978), a compensatory attentional bias toward vision might have caused the tactile stimulus to go unnoticed[4]. It is worth noting at this point that the sense of touch has been suggested to be of great adaptive importance to humans, given that it monitors biologically vital features of the environment by means of direct contact (see Gregory, 1967). That is, tactile stimuli are less likely to be ignored (and hence should be more effective in capturing a participant's attention; e.g., Ho & Spence, 2005; Ho, Reed, & Spence, 2007) because the

---

[4] It should, however, be noted that the relatively lower alerting capability of visual stimuli as compared as tactile stimuli, although asserted by various researchers over the years (e.g., Posner et al. 1976; Von Haller Gilmer 1960), has never, to the best of the knowledge of the author, been put to direct empirical test.

information that they carry is directly available to observers, thus requiring an immediate response (by contrast, Richard Gregory has argued that optical information must be transformed into specific hypotheses regarding the nature of distant stimuli before initialising a response to them). On the basis of such considerations, the results of the present chapter, showing that observers fail to respond to tactile stimuli in favour of visual stimuli, appear to be even more surprising. This apparent contradiction between the claim that tactile stimuli have an inherent capability to capture observers' attention (given their adaptive importance) and the results reported here, showing that people sometimes fail to respond to tactile stimuli when they are presented at the same time as visual stimuli, surely needs to be assessed in future research (see also Santangelo & Spence, 2007, on this point). Perhaps emphasising the embodied nature of the surface on which the tactile stimuli are presented may well make a difference (see Chapter 7).

It should be noted that Ernst & Banks (2002) have put forward an alternative approach to the attentional interpretation of visual dominance, namely the 'maximum likelihood estimation' (MLE) account, as discussed in Chapter 1.1.2. However, the approach is not usually applied to speeded detection tasks. According to these researchers, the nervous system integrates multisensory information in a statistically optimal manner, placing most weight on the modality estimate that shows the least variance (i.e., the most accurate sense) for any given task. Given that under many circumstances, vision is the most accurate sensory modality (especially for tasks that require some kind of spatial discrimination; cf. Kitagawa & Ichihara, 2002), multisensory perception is often dominated and/or shaped by the visual input in situations of multisensory conflict.

The MLE approach has the advantage over Posner et al.'s (1976) account of also being able to provide an explanation for those situations in which vision does not dominate over other sensory modalities (e.g., Guest & Spence, 2003; Shams et al., 2000; however, see Battaglia et al., 2003, for the claim that attention might also play an important, if modest, role within the MLE framework)[5]. However, this approach also assumes that the unisensory components of the multisensory target are redundant and integrated. This specific assumption has been questioned by a number of the researchers who have worked on the Colavita effect (see e.g., Koppen & Spence, 2007a).

The speeded detection task used for the Colavita experiments may involve a high cost function or penalty associated with the multisensory targets. The slower RTs to multisensory targets, coupled with the inhibition of touch are indicative of such a cost function being implemented by the nervous system (see Lockhart & Ting, 2007).

The response latencies for visual stimuli tended to be shorter than those for tactile stimuli in the unisensory trials of Experiments 2.1 – 2.4 (see Tables 2.1, 2.2, & 2.3). Even though this difference in RTs only reached statistical significance in Experiment 2.3, it could nevertheless be argued that more visual-only than tactile-only responses were perhaps made on the multisensory trials because observers responded to the visual stimuli first. However, it should be noted that if observers had first responded only to the visual element of the multisensory stimulus pair, they could still have made a second response to the tactile element should they subsequently have remembered to.

---

[5] Note that the results reported here might also be consistent with the MLE framework if one assumes that visual stimuli simply provide more reliable (i.e., accurate) information about the presence of a stimulus in the environment than do tactile stimuli. In this regard, it is perhaps worth noting that the sense of touch actually performs a dual function, both informing us about events in the environment as well as about the status of our body (see Spence et al., 2001a).

However, the fact that no significant RT differences in the observers responses to the unisensory of the multisensory stimuli were observed on those trials where observers responded correctly for Experiments 2.1, 2.2 and 2.4 argues against this response strategy.

The possibility that the Colavita effect is simply the result of a response bias cannot be conclusively ruled out given the paradigm. However, the results from Experiment 2.4, in which visual input was at threshold, demonstrate an increase in the magnitude of this effect relative to the supra-threshold experimental results (Experiments 2.1, 2.2 and 2.3). Stein et al. 1996 noted that for threshold stimuli, a second stimulus in another modality may act as a priming or 'alerter' signal resulting in response bias. Because the tactile stimuli were perceptually 'matched' in intensity to the 'at threshold' visual input, it is less likely that the inhibition of touch observed is the result of a response bias or alerting function as suggested by Stein et al. (1996).

2.18.1 Spatial modulation of the Colavita effect

The results of Experiment 2.3 revealed that the Colavita visual dominance effect was present no matter whether the visual and tactile stimuli were presented from the same or from different spatial locations, but that the effect was significantly larger when both stimuli were presented from the same position (see Koppen & Spence, 2007b, for similar results in the audiovisual version of the Colavita effect). One possible explanation for these results is that on those multisensory trials in which the visual and tactile stimuli were presented from different spatial locations (sides), they were processed (at least initially) by separate cerebral hemispheres (because they were presented on opposite sides of the body midline, as opposed to on the midline for the stimuli presented in Experiments 2.1, 2.2, & 2.4). It has been suggested elsewhere that the two cerebral hemispheres can be thought of as possessing somewhat discrete pools of attentional resources (e.g., Banich, 1998). Thus, when the visual and tactile stimuli are initially processed in different hemispheres, more cognitive resources should have been available to deal with each of the two stimuli, thus making the tactile stimuli more unlikely to be explicitly reported (cf. Zampini et al., 2003b).

It is, however, important to note that an alternative explanation for why a larger Colavita effect was found when the visual and tactile stimuli were presented from the same location is that the observers may have had less reason to discriminate between the two stimuli, and thus to be aware that two separate events, rather than just one, had been presented. According to research on the 'unity assumption' (e.g., Spence, 2007; Vatakis & Spence, 2007; Welch, 1999; Welch & Warren, 1980), stimuli are increasingly likely to be perceived as reflecting a single, unitary event, the more amodal properties (i.e., those that can be defined without reference to a particular sensory modality, such as size, shape, location, intensity, and temporal characteristics) that they share (Bedford, 2001). As a consequence of multiple stimuli being perceived as reflecting a unitary event, inter-sensory biases between the stimuli increase. In Experiment 2.3, the multisensory stimuli presented from different locations shared their temporal characteristics (i.e., they were presented at the same time and for the same duration) and possibly also their spatial extent (i.e., the light on the observers' fingertip was of roughly the same size as the surface of the vibrating device), but not their spatial location. Therefore, under such circumstances, the two stimuli would presumably have been more likely have been perceived as separate unisensory events, thus perhaps giving rise to a smaller Colavita visual dominance effect. By contrast, when the multisensory stimuli were presented from the same spatial location, they also shared this spatial property. Consequently, the likelihood of unifying the stimuli (and hence any inter-sensory biases consequent on such unification) was increased, thus perhaps explaining why a larger Colavita effect was observed. This might suggest that binding between concurrently-presented characteristics of the stimuli composing a given display may play an important role in the Colavita effect. The increase in the magnitude of the Colavita effect observed in Experiment 2.4 may simply be explained as a consequence of the stimuli having been presented at, or near, threshold.

2.18.2 The Colavita effect and crossmodal extinction

The pattern of observers' responses to multisensory stimuli observed in the present chapter bears a potentially interesting similarity to the clinical condition of crossmodal extinction (e.g., Bender, 1952; Di Pellegrino et al., 1997; see also Vallar et al., 1994, for the neural correlates of visual and tactile extinction)[6]. In particular, patients suffering from crossmodal extinction are often unaware of the occurrence of a stimulus in one sensory modality (e.g., touch) when a stimulus (e.g., vision) is presented at the same time in another modality on the other side of the body/visual field. It has recently been shown that when visual and tactile stimuli are presented simultaneously to patients affected by extinction, their failures to report the contralesional stimuli under conditions of double stimulation are much more pronounced for tactile than for visual

---

[6] It should, however, be noted that for the case of crossmodal extinction, it is the relative position of the two stimuli that determines stimulus is extinguished (i.e., the more contralesional stimulus) while with the Colavita effect, the target modalities of the stimuli determine which stimulus tends to go unnoticed (i.e., the tactile stimulus in the experiments reported here).

stimuli (see Costantini et al., 2007). On the basis of this result, Costantini et al. went on to argue that 'the attentional processing of the visual stimulus inhibits the processing of the tactile stimulus not only across hemispaces but even within the same (contralesional) hemispace), thus indicating that vision dominates over touch in tasks requiring multisensory integration in the space around the body" (see p. 249). Costantini et al.'s results in neuropsychological patients therefore appear to be somewhat similar to the results for the Colavita effect reported here for the first time (in vision & touch) in healthy observers. In particular, both effects seem to be related to the competition between simultaneously presented stimuli (a visual and a tactile stimulus) to gain access to processing resources in the brain.

A variety of different interpretations have been put forward to explain the neurological condition of extinction (in both its unisensory and crossmodal forms). Most researchers now attribute it to damage affecting higher order (rather than peripheral) cognitive functions involved in sustaining a representation of contralesional space or in the deployment of attention to that side of space (e.g., Mattingley et al., 1997; Nico, 1999; Vallar et al., 1994; Ward et al., 1994). This pathological bias affecting spatial representations or attentional systems leads to one stimulus extinguishing the other from awareness when both occur at around the same time and thus compete for processing resources.

It has been claimed by certain researchers that the Colavita effect, and visual dominance effects in general, are due to an imbalance in the access to processing resources; namely between vision and the other sensory modalities (e.g., Colavita & Weisberg, 1979; Posner et al., 1976). Both phenomena would therefore appear to be related to the failure of one stimulus to access awareness when an additional stimulus is presented at the same time. This might be the consequence of a biased competition between the two stimuli for access to processing and/or attentional resources in the brain (e.g., Behrmann et al., 1991; Desimone & Duncan, 1995). That is, the extinguished stimuli (the contralesional stimulus in the neurological condition of extinction and the auditory or tactile stimulus in the Colavita effect) might fail to enter awareness and/or elicit a response because their neural representations elicit an overall lower level of activation in the brain (due to the brain damage, attentional imbalance, and/or other mechanisms).

Recent findings in a patient affected by visual-tactile extinction have shed further light on the neurological correlates of crossmodal extinction in humans (Sarri et al., 2006). Specifically, Sarri and her colleagues reported a neuroimaging study showing that tactile stimuli that have been extinguished from awareness are nevertheless processed unconsciously; and that the competition between the inputs from different sensory modalities in multisensory parietal cortex appears to underlie crossmodal extinction, at least in the one patient that they studied.

Unfortunately, no information regarding the neural correlates of the Colavita visual dominance effect is as yet available. However, a recent review by Spence et al. (2009) provides the first speculations thereof. Moreover, the observation that the effect can be modulated by means of changing the spatial separation between the stimuli (see Experiment 2.3), might lead one to suggest that this phenomenon in neurologically-normal observers, and the extinction effect that is seen in certain neurological patients may be at least, in part, linked by the functioning of spatial processing systems in the brain (such as those involved in directing spatial attention and/or representing space). Moreover, both phenomena might be related to the functioning of those neural structures devoted to the processing, binding, and/or selection of multisensory information in the brain (see Costantini et al., 2007).

The possible parallel between the Colavita effect and the clinical phenomenon of crossmodal extinction (e.g. Bender, 1952; Pellegrino et al., 1976; Sarri et al, 2006) can be seen in relation to the influence of spatial separation of the stimuli in Experiment 2.3. However, another parallel not addressed in this chapter relates to the influence of temporal factors. Temporal factors modulate both the auditory visual Colavita effect (Koppen & Spence, 2007d) and extinction (Cate & Brehmnenn, 2006). Koppen and Spence's study demonstrated that temporal disparity in the auditory and visual components of the multisensory stimulus alters the magnitude of the Colavita effect. They varied the stimulus onset asynchrony between the auditory and visual components of the multisensory targets and found that the effect primarily occurred when the two stimuli fell within the temporal window of audiovisual integration: i.e., when auditory and visual are bound together in to a unified multisensory percept. Moreover, the effect of increased visual dominance with spatial congruency (Experiment 2.3) combined with the temporal influences on audiovisual stimuli show that the Colavita visual dominance effect is largest when the stimuli are perceived as temporally and spatially coincident.

2.19 Conclusions

The results of the four experiments reported in the present Chapter provide the first empirical demonstration that the Colavita visual dominance effect can be found in the visuotactile domain. The dominance of vision over touch reported here might be a consequence of an attentional bias toward the visual modality (e.g., Posner et al., 1976), or to the preferential weighting given to the processing of visual as compared to tactile information by the nervous system (e.g., Ernst & Banks, 2002). This chapter has also shown that the size of the Colavita effect decreases when the visual and tactile stimuli are presented from different positions. Furthermore, the results of Experiment 2.4 demonstrated that the visual dominance over touch in the Colavita task still occurs when the intensity of the two stimuli have been matched psychophysically, and the magnitude of the effect increases when the stimuli are at or near threshold. It has been argued that the Colavita effect may bear interesting similarities to the neuropathological syndrome of crossmodal extinction. Future research should further examine the neural and cognitive correlates of the Colavita effect using brain imaging techniques and signal detection theory. Furthermore, it will also be interesting to investigate in future research whether the visual dominance over touch is reduced for visual stimuli that are presented 'off' as compared to 'on' a participant's body (cf. Gallace et al., 2006; Hari & Jousmäki, 1996). This embodied component of the Colavita effect will be explored in the following chapter.

Chapter 3: Extending the visual dominance over touch for input off the body

3.1 Introduction

In Chapter 2, the Colavita paradigm was used to investigate the nature of any crossmodal interactions between complementary visual and touch cues in a speeded detection task. The concurrent presentation of a visual stimulus was found to inhibit observers' responses to touch (with a higher probability than it itself was inhibited) when observers had to respond as quickly and accurately as possible to both elements of the multisensory visuotactile cues. This crossmodal "extinction" of touch by vision even occurred when the subjective intensity of the visual and tactile stimuli had been matched (cf. Experiment 2.4). The two complementary cues appeared to be in competition for representation in the participant's awareness. Yet the fact that vision should extinguish the perception of touch is in slight contradiction to work from an evolutionary perspective that suggests that tactile stimuli ought to be of great adaptive importance to humans since tactile stimulation normally signals direct contact between external stimuli and one's body surface (Gregory, 1967). Moreover, it has been argued that tactile input may sometimes signal a potentially dangerous situation requiring a rapid response on the part of the organism). Visuotactile interactions seem to be strongest on the body (in personal space), and in the area directly surrounding it (in peripersonal space; e.g., Cardini, Longo, Driver, & Haggard, in press; Jacobs, Brozzolli, & Farné, 2012; Spence, Pavani, Maravita, & Holmes, 2008). Neurophysiological support for this view has come from recent studies of patients suffering from tactile extinction resulting from right-hemisphere brain damage. Researchers have demonstrated that visual stimuli presented proximal to a patient's ipsilesional hand can extinguish their awareness of a concurrent tactile stimulus delivered to the contralesional hand. However, presenting the visual stimulus far from the patient’s ipsilesional hand has been shown to ameliorate this form of crossmodal extinction (Borgers, Hauser, & Himmelbach, 2011; di Pellegrino et al., 1997; Farné & Làdavas, 2000; Làdavas, di Pellegrino, & Zeloni, 1998; Lane, Ball, Smith, Schenk, & Ellison, 2011). It should be noted that tactile, as opposed to auditory events occur on the body surface and are therefore body centred (see Hanson et al., 2009).

The finding that visuotactile crossmodal effects tend to be more pronounced in the space immediately surrounding the body has been suggested to arise from an integrated visuotactile map of peripersonal space, based on body-centred coordinates (Dufour, & Touzalin, 2008; Farné et al., 2000; Graziano & Gross, 1998; Mattingley et al., 1997). To date, evidence for such a map has come primarily from animal studies (Graziano & Gross, 1994; Graziano & Gross, 1998; Makin, Holmes, & Ehrsson, 2008), although Farné et al. 2000 have demonstrated the first behavioural evidence for this in the sub-population of right brain-damaged extinction patients. Given the behavioural similarities between the Colavita visual dominance effect and the phenomenon of crossmodal extinction (see Egeth & Sager, 1977; Farné et al., 2000; Hartcher-O’Brien et al., 2008; Koppen & Spence, 2007c) and the fact that changes in the mapping between vision and touch depend on the region of space (personal or peripersonal) in which they are presented, the current chapter combines these observations in order to try to clarify the possible mechanisms underlying both the visual dominance effect and its relation to body-centred spatial representations.

A paradigm that is often used in order to investigate multisensory integration as it relates to body representation is that of the rubber hand illusion (Botvinick & Cohen, 1998; Ehrsson et al., 2004; Farné et al., 2000; Holmes, Snijders, & Spence, 2006). If a rubber hand is placed in a plausible position relative to a participant's own (occluded) body part, visual stimulation of the rubber hand coupled with tactile stimulation of the participant's own hand produces the impression that the tactile sensation is coming from the rubber hand and that the visible rubber hand is the participant's own (Botvinick & Cohen, 1998). This illusion is decreased if the artificial limb is placed such that it is inconsistent with the real position and/or if visual and tactile stimulation are asynchronous. Moreover, Holmes et al. (2006) demonstrated that recalibration of personal space can occur even without stroking the participant's own hand. In their study, they found a shift in perceived location of the real hand, towards the rubber hand. Furthermore, they found that this effect is not dependent on perceived ownership of the rubber hand – when given the same questionnaire used by Botvinick and Cohen (1998), observers did not report feeling as if the rubber hand were their own. The rubber hand manipulation (the inclusion of an artificial limb in a plausible position relative to the subjects own body in the experimental setup) therefore provides an elegant way of interrogating how vision and touch interact in the space directly surrounding the participant's body, and in certain cases, independent of establishing perceived ownership of the artificial limb.

Farné et al. (2000) observed that when the visual stimuli were presented on a distal rubber hand, their patients' perception of touch was extinguished to the same extent as it had been when the visual stimuli were presented on the patients' own bodies. Given this result, and the claimed similarity between extinction and the Colavita visual dominance effect, the current prediction for the behavioural outcome in this chapter is that the on- versus off- body manipulation would mediate the visual dominance effect in neurologically-normal individuals. That is, it is expected that visual information presented on the body, or on a rubber hand, will dominate over touch, but that visual information off the body would have less of an effect upon the detection of tactile stimuli. Given Holmes et al.'s (2006) results, no explicit attempt was made to create a conscious perception of ownership of the rubber hand; it was expected that the mere presence of the rubber hand at a plausible position would be sufficient to induce the visual dominance effect.

Observers had to make speeded detection/discrimination responses to a random sequence of visual-only (presented on- or off-body), tactile-only, and visuotactile cues. The results for the tactile, the visual (presented on- or off-body), and multisensory (presented on- or off-body) responses are presented below.

3.2 Method

3.2.1 Observers

Twenty-two observers (10 male), all right-handed and aged from 19 to 36 years (mean age 26.2 years) took part in the rubber hand absent condition (RH-absent). One participant who failed to respond on more than 75% of the trials was removed from the dataset. Twenty-one naïve, right-handed observers (9 male), aged between 19 and 35 years (mean age 25.7 years), took part in the rubber hand present condition (RH-present). All the observers had normal or corrected-to-normal vision. All of the observers were naïve as to the purpose of the experiment. The experiment was conducted in accordance with the ethical guidelines laid down in the Declaration of Helsinki. The experiment took approximately 30 minutes to complete and the observers received of a 5 pound (U.K. Sterling) gift voucher in return for their participation.

3.2.2 Apparatus and materials

Each participant sat on a chair in a dimly-illuminated testing booth, with their left index finger inserted into a custom-built stimulus box, positioned on a table approximately 65 cm in front of their torso, 65 cm to the left of their body midline (see Figure 3.1). The visual stimulus consisted of a yellow light emitting diode (LED), CIE (0.4, 0.47), with a luminance of 0.98 cd/m2, presented for 50 ms. The LED was mounted in the top of the black stimulus box pointing downward. It was positioned so as to illuminate the tip of the participant's left index finger from above. The off-body visual stimulus was presented 65 cm to the right of the participant's midline. The off-body LED was also mounted in a black box similar to the one used to deliver the on-body stimuli and was orientated in a similar manner. The interior of the stimulus boxes used to present the stimuli had a matte black finish to avoid any unwanted reflection from the LED.

A rubber hand was placed directly below the off-body visual stimulus in the RH-present condition (Figure 3.1B). The rubber hand constituted a plausible representation of a male human hand. The hand was 40 cm long, 5 cm at the narrowest point, and 8 cm wide at its widest point. The orientation of the rubber hand mirrored that of the participant's left hand. The LED illuminated the fingertip of the rubber hand in a manner similar to the illumination of the participant's left hand. In both conditions, the participant rested their right hand in a relaxed position by their side. The on- and off-body visual stimuli were presented equidistant from the participant's midline. The tactile stimulus was presented by means of a bone conduction vibrator (Oticon Limited, B/C 2-PIN, 100 Ohm, Hamilton, UK) with a 1.6 x 2.4 cm vibrating surface. The 250 Hz supra-threshold vibrotactile stimulus was generated by a Taylor TE-22 Sine-Square Wave audio signal generator (Dynacomp, Inc. Fayette, NY), and presented at a verified, clearly perceivable level. The vibrator was mounted in a foam block placed directly below the on-body LED in the custom-built stimulus box. The participant was instructed to place his/her left index finger directly over the vibrator, thus ensuring that the on-body visual stimulus and the tactile stimulus were presented from exactly the same spatial position.

Speeded detection/discrimination responses were collected using two foot-pedals placed on the floor below the table (one placed 10 cm to the left and the other 10 cm to the right of the participant's midline), on which the participant rested their feet. The participant was instructed to lift his/her toes off of one pedal in response to the visual cues and to lift the toes of his/her other foot off of the other pedal in response to the presentation of the tactile cues. The allocation of the stimuli to the response pedals was counterbalanced across observers. The observers were explicitly instructed to respond to the multisensory cue trials by lifting their feet off both foot pedals, though no specific instructions were provided as to whether they should lift the two foot-pedals simultaneously or not. White noise was presented continuously at 55dB SPL over closed-ear headphones in order to mask any background noise (including the noise generated by the activation of the vibrotactile stimulator).

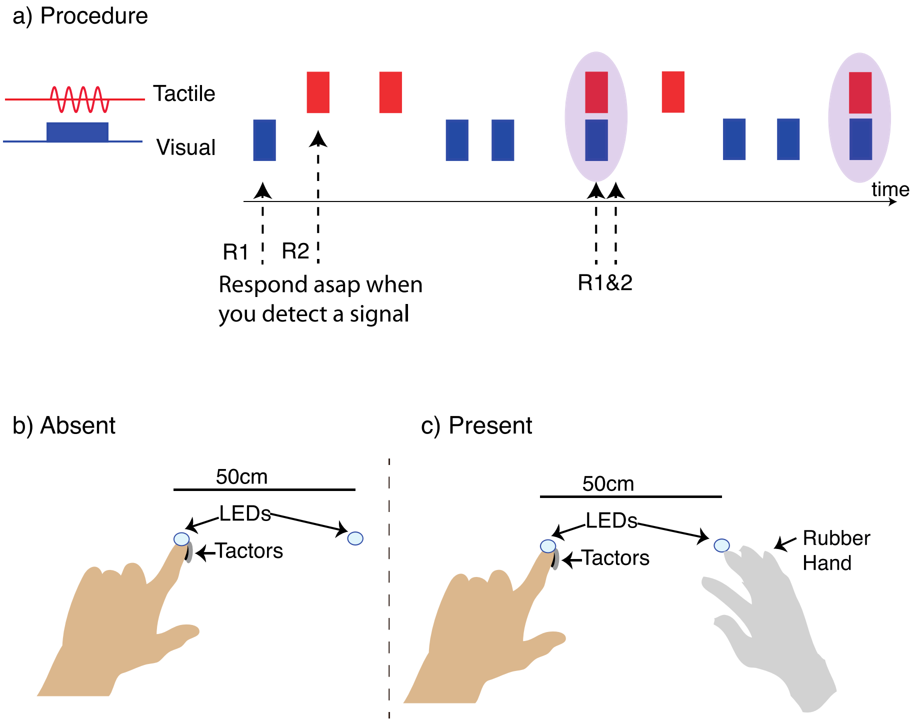


Figure 3.1. Schematic outline of experimental set-up for the Colavita paradigm with visual stimuli presented both on and off the observers body. a) The procedure which is identical to that presented in Chapter 2.2.3. c) The Rubber hand-absent condition in which the off-body visual stimulus was presented in peripersonal space (50cm from the observers body) and b) the RH-present condition in which a plausibly-positioned rubber hand was placed in the location of the off-body visual stimulus from the RH-absent condition. R1 and R2 indicate the choice responses to visual and tactile input.

3.2.3 Design

Each participant completed 6 blocks of 100 trials. Each block of trials consisted of 40 unisensory visual (20 unisensory visual-on body, 20 unisensory visual-off body), 40 unisensory tactile, 20 multisensory (10 multisensory visual-on body, and 10 multisensory-visual-off body) with the same probability as in Experiments 2.1, 2.3 and 2.4. Note that this ratio of multisensory to unisensory trials is standard for research on the Colavita task (e.g., see Koppen, Alsius, & Spence, 2008; Koppen & Spence, 2007a, b). For the RH-present condition, the right index finger of the rubber hand was placed at the off-body location. The order of presentation of the stimuli was randomised within each block of experimental trials. A block of 30 practice trials was presented prior to the main experimental session. These trials were identical to the experimental trials but were not analysed. A between-participants design was used, as to not further draw observers' attention to the experimental manipulation of the presence vs. absence of the rubber hand.

3.2.4 Procedure

The participant was instructed to fixate the centrally-positioned fixation cross on an otherwise blank screen. The participant was presented with a unisensory visual (presented on- or off-body), unisensory tactile, or multisensory cue (once again with the visual stimulus presented on- or off-body), on each trial. The stimuli were presented for 50 ms followed by a 1500 ms response interval, after which time the next trial began. The observers were instructed to respond as rapidly and accurately as possible following the presentation of the cue. No feedback was provided regarding the correctness of a participant's responses.

The observers in the RH-present condition had to complete a RHI questionnaire (see Botvinick & Cohen, 1998) after having completed the behavioural task, in order to measure any sense of ownership of the rubber hand that they may have experienced. It should, however, be noted that no specific attempt was made to establish ownership of the rubber hand by, for example, simultaneously stroking the rubber hand and the participant's own right hand (cf. Botvinick & Cohen, 1998). Responses to the questionnaire items were therefore not expected to necessarily reveal any sense of ownership in the present chapter (although see Holmes et al., 2006).

Table 3.1. Mean unisensory and multisensory error rates (%) and RT (ms) for the RH-absent and RH-present conditions. Performance measures (i.e., not Colavita effect magnitudes – which reflect difference scores) are presented (in bold font) with standard error of the mean (SEM) in brackets.

	Unisensory cue			Multisensory cues			
	Tactile	Visual on-body	Visual off-body	Visual on-body		Visual off-body	
				Tactile-only	Visual-only	Tactile-only	Visual-only
RH-absent							
Error rates (%)	5.8 (1.6)	3.6 (1.1)	4.4 (0.9)	4.7 (0.7)	19.5 (3.3)	4.7 (0.7)	19.5 (3.3)
Total	**5.8 (1.6)**	**3.6 (1.1)**	**4.4 (0.9)**	**22.2 (3.4)**		**4.8 (2.9)**	
Reaction times (ms)	422 (187)	573 (255)	614 (154)	422 (187)	643 (134)	422 (187)	643 (134)
RH-present							
Error rates (%)	4.5 (0.8)	6.7 (1.1)	4.7 (1.4)	1.3 (0.4)	4.2 (0.7)	1.3 (0.5)	8.1 (1.8)
Total	**4.5 (0.8)**	**6.7 (1.1)**	**4.7 (1.4)**	**5.3 (1.1)**		**9.4(1.3)**	
Reaction times (ms)	537 (175)	473 (208)	554 (188)	591 (188)	601 (208)	577 (161)	554 (181)

3.3 Results

On average, the observers failed to make any responses on 2% of the trials (similar to the number reported in Chapter 2). These trials were not included in the subsequent data analyses. Failure to respond to a unisensory cue was treated as an omission for the purpose of the analyses. The results are summarised in Table 3.1. Observers made numerically slightly more unisensory tactile errors than unisensory visual errors in the RH-absent (no rubber hand was located in the space of visual stimulation off the body) and more unisensory visual-on body than tactile unisensory in RH-present conditions though this difference was not significant [$F<1$, n.s.].

3.3.1 Error data

The magnitude of the visual dominance effect was calculated for the RH-present and RH-absent conditions using the multisensory trials in which the observers only responded to one of the two stimuli. The magnitude of the visual dominance effect was quantified as the number of visual-only minus the number of tactile-only responses (see Figure 3.2). In order to determine whether or not there was a difference in the magnitude of the visual dominance effect between the RH-present and RH-absent conditions, an (ANOVA) was conducted on the data with the between-participants factor of Condition (RH-absent vs. RH-present) and the within-participants factors of Response type (visual-only vs. tactile-only) and Position of the visual stimulus (on vs. off of the participant's body). The analysis of the error data from the multisensory cue trials revealed a main effect of Response type [$F(1, 40)=13.83$, $p<0.001$, $\eta^2=0.25$], with observers making significantly more visual-only ($M=8.6\%$) than tactile-only responses ($M=3.2\%$). There was also a significant main effect of the Position of visual stimulus [$F(1, 40)=28.72$, $p<0.001$, $\eta^2=0.48$], with observers making significantly more errors when the visual stimuli were presented on ($M=10.1\%$) as opposed to off their body ($M=2.3\%$). Moreover, there were also significant interactions between Response type and Condition [$F(1, 40)=18.72$, $p<0.001$,

η^2=0.32], between the Position of the visual stimulus and Response type [F(1, 40)=6.61, p=0.02, η^2=0.13], and between Condition, Response type, and Position of the visual stimulus [F(1, 40)=7.58, p<0.01, η^2=0.16]. Looking further into these interactions, Bonferroni-corrected post-hoc comparisons revealed that the visual dominance effect was significantly larger when the visual stimuli were presented on the participant's body than when they were not in the RH-absent condition [t(41)=3.29 p<0.01]. Bonferroni-corrected post-hoc comparisons also revealed a significantly larger visual dominance effect when visual stimuli were presented on the body in the RH-absent condition than for the same location in the RH-present condition [t(41)=2.31, p<0.05].

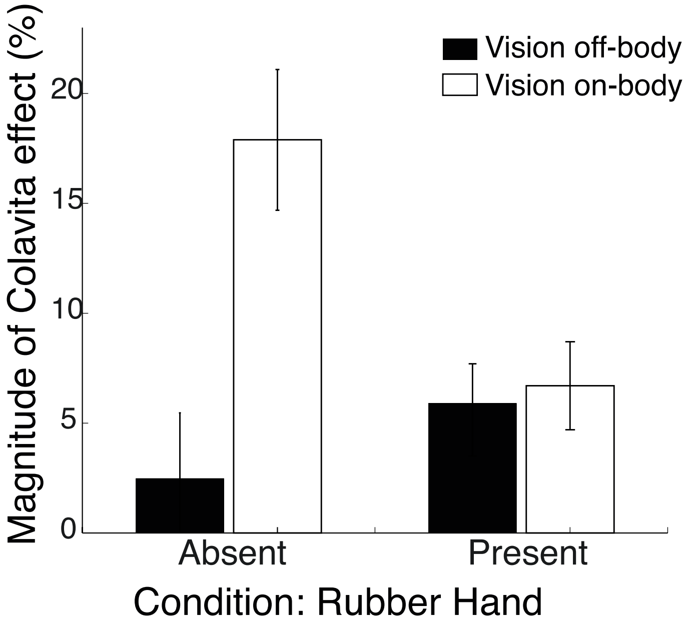


Figure 3.2. The magnitude of the visual dominance effect (calculated as the number of visual-only minus tactile-only responses on the multisensory cue trials) for the RH-absent and RH-present conditions. Error bars represent the standard errors of the means.

3.3.2 RT data

The mean RTs for those trials in which the observers responded correctly were also analysed using a mixed-model ANOVA with the factors of Response (visual-on, visual-off, or tactile), Trial Type (unisensory vs. multisensory), and Rubber Hand (present vs. absent). This analysis revealed a significant main effect of Trial Type [F(1, 19)=15.78, p<0.01, η^2=0.45], with observers responding more rapidly to the unisensory visual cues (M=509 ms) than to multisensory tactile cues (M=588 ms) as verified by post-hoc Bonferroni-corrected comparisons [t(20)=2.56, p=0.01]. However, the main effect of Response was not significant [F<1, n.s.], nor was the between-participants factor of Rubber Hand (present vs. absent) [F<1, n.s.].

3.3.3 Rubber Hand Illusion Questionnaire

The items assessing body ownership from the questionnaire were analysed using one-sample one-tailed t-tests in order to verify whether observers' experienced a RH illusion (Botvinick & Cohen, 1998). In the one-sample t-test, mean responses were compared to a score of 4 (indicating 0 or neutral on the Botvinick & Cohen questionnaire), which would indicate no effect (just as in Botvinick and Cohen's original questionnaire). Responses to Question 3 (experience of ownership of the rubber hand) revealed that the observers did not experience any sense of ownership of the rubber limb [t(20)=-3.08, p<0.01]. Question 1 investigated a possible perceived shift in the location of the tactile stimulus to the position of the rubber hand. A similar one-tailed one sample t-test analysis revealed that the tactile stimulus was not perceived as originating from the rubber hand [t(20)=-5.57, p<0.001]. These results therefore suggest that the observers in the RH-present condition did not experience ownership (Q3), nor was the touch sensation ventriloquized and perceived at the location of the rubber hand (Q1).

3.4 Discussion

The results of the experiment reported in this chapter demonstrate a robust visual dominance effect over touch, thus replicating the findings reported in Chapter 2. Furthermore, the current results also show that the dominance of vision over touch was significantly reduced when the visual stimuli were presented away from (rather than on) the participant's own body in line with predictions. This result, obtained in neurologically-normal observers, is consistent with previous neuropsychological data showing that presenting the visual component of a visuotactile cue on a (right-hemisphere brain-damaged) patient's body reduces their ability to perceive tactile stimuli presented to the contralesional hand (see Làdavas et al., 2000), while increasing the distance of the visual stimulus from a patient's body weakens its dominance over touch (Farné et al., 2000). The key finding to emerge from the rubber hand manipulation incorporated into the present chapter was that the visual stimulus 'extinguished' normal observers' perception of touch to the same extent when presented to their fingertip or on the fingertip of a rubber hand. That is, visuotactile interactions on a real or artificial body part diminish the ability of neurologically-normal observers to respond to touch. Presenting the same stimuli to the same region of space with no limb present does not have the same effect. This finding is consistent with the suggestion (based on the neuropsychological literature) that visuotactile interactions occur in a body-centred frame of reference (see Farné et al., 2000; Valenza, Murray, Ptak, & Viulleumier, 2004). It is, however, important to note that the on- vs. off-body manipulation of the location of the visual cues utilised in the present chapter was perfectly confounded with the manipulation of whether the visual and tactile stimuli were presented from the same spatial location or not. As such, it is important to point out here that the on- vs. off-body results contrast with those reported in Chapter 2 (Experiment 2.3), in which a similar manipulation of same- versus different position multisensory trials was incorporated. In that experiment, the visual stimuli were always presented to one of the observers' hands (with one hand placed 12.5cm on either side of midline). The visual dominance effect reported there was still significant, albeit slightly reduced in magnitude, when the stimuli were presented from different locations. The results from Chapter 2 therefore suggest that the difference in the magnitude of the visual dominance effect was not due to stimuli coming from separate spatial locations. Thus, the lack of any visual dominance off the body in the RH-absent condition of the present chapter cannot simply be explained in terms of the spatial incongruence between vision and touch. Furthermore, the spatial separation was equivalent in both the RH-absent and RH-present conditions, while the magnitude of the visual dominance effect differed dramatically between the two conditions.

In Chapter 2, the similarities between crossmodal tactile extinction and the Colavita effect reported in that chapter were outlined. Additionally, research suggests that for extinction patients (Farné et al., 2000; Ladavas, et al., 1998), moving the visual stimulus away from – or off – the patient's body, decreases visual dominance over touch (see di Pellegrino et al., 1997a; Farné & Làdavas, 2000; Làdavas et al., 1998). These results coupled with those presented in the current chapter, suggest a representation of space based in limb-centred coordinates that is sensitive to competition between multiple sensory cues (Valenza et al., 2004). However, Valenza et al., (2004) note that further research has to be done to disentangle the contributions of limb-centred and space-centred representations to visual-tactile discrimination. The premise motivating the current experiment was that visuotactile interactions are most pronounced in the space directly surrounding the body using a very simply paradigm with neurologically-normal observers (see Dufour & Touzalin, 2008; Farné et al., 2000; Graziano & Gross, 1998). The observed decrease (and the subsequent re-instatement of visual dominance when the light occurred on a comparably-positioned rubber hand) is consistent with this, and thus with the hypothesis that the coding of visuotactile peripersonal space is based on limb-centred coordinates (Graziano, & Gross, 1998; see also, Makin, Holmes, & Zohary, 2007). Furthermore, the existence of such a limb-centred coding system is further supported by the current findings in neurologically normal observers.

An important difference between the current experiment and that of Farné et al. (2000) is the population used to explore the sensory interactions. The major contributing factor in crossmodal tactile extinction in patients is the lesion in the right hemisphere. In drawing an analogy between visuotactile extinction and the Colavita effect it is not easy to link both behavioural phenomena to a similar mechanism. However, both the Colavita visual dominance effect (Colavita, 1974) and crossmodal extinction (Inhoff, Rafal, & Posner, 1992; Mattingley et al., 1997) have been related to attention (see Spence, 2009 for a review). Given a similar visual dominance with the light presented to a distal rubber hand in the current chapter and in the data for right brain damaged patients demonstrating crossmodal extinction, the current findings create further parallels between both phenomena.

Despite the fact that the rubber hand was not incorporated into the conscious body representation (as indicated by responses to the RHI questionnaire; see Holmes & Spence, 2006), the presentation of the visual stimulus to the fingertip of the rubber hand was sufficient to maintain the dominance of vision over touch. Moreover, Holmes et al. (2006) did not attempt to establish the illusion via stroking, but also observed a dissociation between observers' responses on the RHI questionnaire and changes in behavioural responses. In particular, they found that observers' reaching behaviour was affected by visual exposure to a (mirror displaced) real hand or to a rubber hand, even though the RHI questionnaire results demonstrated no sense of ownership (nor did they find a strong relationship between RHI responses and reaching behaviour).

The similarity between the Holmes et al. (2006) findings and the results of the RHI questionnaire reported here supports the notion that, a) the correlation between RHI Questionnaire and other behavioural measures of ownership (i.e. proprioceptive drift or visual dominance as shown here) is not always positive (Holmes et al., 2006) and that b) in the present chapter, observers need not have been aware of a recalibration of personal space (e.g. Holmes et al., 2006; Kammers et al., 2009), but nonetheless updated their representations to incorporate the rubber hand. This may suggest that the implicit encoding of the presence of an artificial limb in peripersonal space is sufficient to activate the multisensory coding system (cf. Holmes et al., 2006). That is, the rubber-hand illusion is not in and of itself a necessary precondition for remapping to occur; it may be that when the illusion does occur, it is a consequence of remapping rather than the cause.

Another possibility is that the questionnaire was not sufficiently sensitive to detect a sense of ownership, which may have simply been present to a lesser extent than in the original Botvinick and Cohen (1998) study. Alternatively, however, the presence of the rubber hand may have resulted in the division of observers' attention, equally anchored to real and artificial body parts, therefore reducing the size of the visual dominance effect for the on-body location, but increasing it slightly in the off-body position (see Abrams, Davoli, Du, Knapp, & Paull, 2008; Hari & Jousmäki, 1996; Reed, Grubb, & Steele, 2005). This interpretation offers tentative support for the bias in attentional resources (preferentially toward vision) explanation of the visual dominance effect (cf. Egeth & Sager, 1977; Hecht & Reiner, 2009; Koppen & Spence, 2007a).

RT responses to the different signal conditions demonstrate that observers are faster to respond to the tactile component generally, but for visual signals presented off the body on the artificial limb, responses are faster than to a coupled tactile component. Such trends are aligned with the idea that the presence of the rubber hand changes the attentional capacity of the visual stimulus (as indicated by a faster response time). In a model of competition for resources (Sinnett et al., 2008), this would result in the tactile taking precedence with more tactile responses to multisensory cues. Given that this dominance of tactile responses to multisensory cues overall was not found, the competition for resources strategy does not explain the results without calling on further processing constraints. Moreover, RT data are sampled from an inherently noisy and non-normal distribution (e.g., Carpenter & Williams, 1995) and should therefore be treated with caution.

Disentangling the role of sensory versus cognitive and decisional factors is inevitably complex. Previous work on the Colavita visual dominance effect with audiovisual cues (Koppen, Levitan, & Spence, 2009) explored this question using signal detection theory and found that observers displayed a slight but significant reduction in sensitivity to auditory stimuli in the presence of vision, suggesting that the effect is, at least in part, perceptual rather than decisional. For visuotactile stimuli, it is unclear whether the findings would correspond directly, but it is expected that if observers were solely responding based upon demand characteristics of the experiment, this would impact their responses to the questionnaire. Given that they do not, it is possible that a signal detection analysis of visuotactile interactions would show a decrease in tactile sensitivity, consistent with the phenomenon of tactile extinction.

The visual dominance over touch reported in this chapter supports the argument that when visual stimuli are presented near the hand they recruit multisensory neurons (e.g., Farné et al., 2000; Gross & Graziano, 1998). In so doing, these neurons facilitate the detection of a cue when the visual stimulus is anchored on a body-part relative to a non-limb anchored visual stimulus (see Kao & Goodale, 2008). The results of the RH-absent condition are consistent with this claim. The results of the RH-present condition indicate that it is not necessary for the visual stimulus to illuminate a real body part; a plausible representation of a body part is sufficient to elicit the effect as well. With regard to the latter point, it is worth noting that Igarashi, Kimura, Spence, and Ichihara (2007) have shown that even simply viewing a picture of a hand can give rise to alteration (inhibition/facilitation) in observers' tactile discrimination responses.

Visuotactile interactions are therefore strongest in association with body parts, real or artificial.

The findings presented in Chapter 2 when taken together with the findings reported in the current chapter point to an instance of multisensory complementary cue interactions in which one cue wins in competition for representation. Why vision? Many theories of visual dominance from attention (Duncan et al., 1997) to sensory acuity (Bertelson & De Gelder, 2003; Ascherleben & Bertelson, 2003) and brain areas devoted to visual processing (Duncan et al., 1997) try to illuminate why this is so. These possibilities are elaborated further in Chapter 8. The findings in this chapter demonstrate that visual dominance occurs when observers are required to respond as quickly as possible to the cues. This response appears to incorporate a lengthy decision stage evident in the slow RTs. Furthermore this effect is dependent on some body-centred contingency (see Spence et al., 2012 for a review).

3.5 Conclusions

The different cues in the environment provide the nervous system with different types of cues depending on the task and goal of the participant. While observers' responses were highly accurate for unisensory cues, their performance with the visuotactile cues was deleteriously affected by the simultaneous-presentation of a visual stimulus on (as opposed to away from) their body. In particular, when the visual stimulus was presented to their fingertip, observers failed to respond to far more of the tactile than visual stimuli on multisensory trials. The magnitude of this visual dominance effect decreased significantly when the visual stimulus was moved off the body. When a rubber hand was placed at the off-body location, a similar (albeit reduced) visual dominance effect was observed in both positions. These results therefore suggest that visuotactile interactions are strongest when visual stimuli are presented on a body (no matter whom that body, or body-part, belongs to).

3.5.1 Interim summary

While both this and the previous chapter have focused on the more spatial influences on the visuotactile Colavita effect, one possible reason for the effects occurrence, as suggested in Chapter 1, is the statistics of the environment in which the nervous system finds itself for the Colavita experiments. What are the statistics of the environment? The Colavita effect occurs most strongly when the number of multisensory trials is limited (see Colavita, 1974; Koppen & Spence, 2007c; Chapter 2 of this thesis Experiment 2.2). When the probability of these multisensory trials increases, the effect decreases. When multisensory trials are increased to 90% of the total number of trials in an experiment, the effect disappears altogether (Koppen & Spence, 2008). This aspect of environmental short-term environmental predictions, coupled with the nature of the cues – i.e. complementary, due to the task, could account for the Colavita inhibitory response effect.

In order to explicitly test the influence of temporal sequences on subsequent speeded detection responses, the following two chapters explore an intriguing phenomenon known as temporal recalibration (Fujisaki et al., 2004), but for audiovisual as opposed to visuo-tactile stimulus combinations.

Chapter 4: Adaptation to audiovisual asynchrony modulates the speed with which sounds are detected

4.1 Introduction

To build a coherent and robust representation of the multisensory environment in which we live, we need to be able to perceive the redundant multisensory aspects of a physical event as occurring simultaneously. In order to do this, the nervous system must be able to deal with physical and neural differences in delay occurring across the different modalities (see Section 1.2.2). Several potential mechanisms have been put forward in order to account for how multisensory asynchronies are dealt with by the nervous system (Keetels & Vroomen, 2010; King, 2005; Spence & Squire, 2003; Vroomen & Keetels, 2010). In this chapter, as introduced in Section 1.2.2, the effects of continuous exposure to asynchrony in redundant signals will be explored via the phenomenon of temporal recalibration.

Consider a live television broadcast where a reporter transmits from a remote location. Technical limitations with the television apparatus can cause the voice of the reporter to be desynchronised with the visual stream, usually meaning that the audio signal leads. At first, the relative delay between the two redundant streams of information creates the impression that the voice and the moving face of the reporter are emanating from two different individuals. Only after some time, by experiencing the correlation between the voice steam and mouth movements, can the nervous system realize that there is a correspondence between the two streams of information and therefore deal with the temporal inconsistency (Helbig & Ernst, 2007; Körding et al., 2007; Parise et al., 2011). As you keep watching the TV show, while being exposed to the asynchrony, the temporal offset becomes less noticeable perceptually. Even though one is still aware of the temporal discrepancy, the audiovisual information about the reporter becomes unified into one coherent speech percept it is possible to enjoy the broadcasted show.

Recent studies have demonstrated that the nervous system can adjust the processing of asynchronous sensory signals in order to help preserve the subjective impression of simultaneity. Prolonged exposure to redundant yet asynchronous stimuli (such as 'simple' pairs of beeps and flashes, or 'complex' audiovisual speech) often induces temporal 'aftereffects' (see Section 1.2.2.1) in the perception of subsequently presented stimuli (see Di Luca et al., 2009; Fujisaki et al., 2004; Hanson et al., 2008b; Harrar & Harris, 2008; Keetels & Vroomen, 2010). The mere presence of these temporal "aftereffects" suggests that the mechanisms for integrating information across the sensory channels are flexible in terms of reducing temporal disparities and hence optimizing the perception of the objects and events around us. Crucially, Fujisaki et al. (2004) provided (indirect) evidence of temporal aftereffects even in perceptual illusions that for their occurrence depend on the temporal relation between signals, perhaps suggesting that these effects are genuinely perceptual.

It has further been suggested that these temporal recalibration effects might be based on a readjustment of the speed at which different sensory signals are transmitted neurally (e.g., Fujisaki et al., 2004; Vroomen et al., 2004). However, previous research has not been able to clarify whether or not this readjustment results in a genuine change in the speed at which humans respond to either of the recalibrated stimuli.

Audition has traditionally been seen as a sensory modality that dominates over the others (e.g., vision or touch) for temporal estimates, thus resulting in illusory effects such as the so-called temporal ventriloquism effect (see Hartcher-O'Brien & Alais, 2011; Morein-Zamir, Soto-Faraco, & Kingston, 2003; Welch, DuttonHurt, & Warren, 1986). However, as will be demonstrated in Chapter 7, auditory driving can be reversed when the auditory signal is degraded. Yet disregarding the case where noise has been added to the signals, the nervous system has ample experience of auditory processing being more precise for temporal properties of the environment. Following on from this idea, the mechanism underlying audiovisual temporal adaptation could reasonably be expected to imply a shift in the processing of the visual signal toward the (perhaps more accurate) auditory signal (see "Predicted after-effects" in Figure 4.1). It is, however, worth highlighting the fact that the most likely cause of audiovisual asynchrony in our environment is the physical delay between auditory and visual signals induced by the fact that light travels much faster than sound (nearly 300,000,000m/s vs. 340m/s in the air, respectively; see Fain, 2003). When taken together with differences in the neural transduction of sensory signals at the sensory epithelia, it has been estimated that sounds will reach primary cortex before visual signals for events that occur up to 10m away from us, but that visual signals will lead whenever an event occurs at a greater distance (see Pöppel, 1997; Spence & Squire, 2003). Given that only auditory arrival time is affected by a change in event distance, visual information actually provides a far more precise estimate of when an event occurred. Keeping this fact in mind, it makes sense for the nervous system to 'pull' auditory signals into temporal alignment with the corresponding visual events and not vice versa (see "Predicted after-effects" in Figure 4.1). The question of which signal (auditory, visual, or both) is modulated during the adaptation to asynchronous audiovisual signals is, then, by no means a trivial one.

In the experiment reported in this chapter, simple RTs to individually presented auditory and visual stimuli were used to investigate whether the temporal readjustment observed during (and after) the exposure to asynchronous sensory signals influences the speed at which the nervous system can respond to unimodal sensory stimuli. Interestingly, in a recent study by Harrar and Harris (2008), RTs were used in a control experiment. Their recording of RTs was implemented in order to investigate the possible influence of attention to the asynchronous stimuli on the typical shifts observed in psychophysical tasks such as SJs. Harrar and Harris' study (Harrar & Harris, 2008) represents an interesting attempt to investigate the effects of perceiving audiovisual asynchrony on the detection of visual or auditory stimuli. However, while the methodology used by these authors was designed to investigate any possible effects of attention, it did not allow them to observe specific modulations of the RTs to visual and auditory stimuli due to temporal recalibration. In contrast with this previous study, there were two different kinds of exposure in the current experiment: exposure to synchrony and exposure to asynchrony (not just to asynchrony, see Harrar & Harris, 2008). Moreover, in the current experimental design (see Section 4.2.2), two different kinds of asynchrony (VA and AV) were used instead of just one. Moreover, less RT trials (4 as opposed to 12) were presented between re-exposure top-ups (see Section 4.2.3), thus reducing the possible impact of practice/fatigue or temporal readjustment to baseline. From this it is clear that the current findings are due to the prior exposure to asynchrony, and not to any other possible factor.

Regarding the use of RTs and psychophysical measures to study the temporal aspects of human perception, it is still not clear whether the two approaches address the same nervous system processes or not (see Jaskowski, 1996; van Eijk, Kohlrausch, & Juola, 2008). It has recently been suggested that, although obvious differences arise between TOJs and RTs at a level where decisions are made, the two tasks seem to be based on the same system at early stages of processing (Cardoso-Leite, Gorea, & Mamassian, 2007). This evidence would perhaps justify the use of RTs to directly assess temporal recalibration between sensory signals.

In the present chapter, RTs after exposure to asynchrony (either audition lagging (VA) or audition leading (AV)) were compared, for the first time, with RTs after exposure to synchronous audiovisual stimulation (see Figure 4.1). According to the “temporal shift” hypothesis, which claims that the speed of signal processing is modulated in order to preserve the perception of simultaneity, one should expect to find very specific modulations of observers’ RTs depending on the direction of the exposure to asynchrony (e.g., faster RTs to sound, but slower RTs to lights, would be seen following exposure to VA asynchrony). Surprisingly, the results reported in this chapter suggest that only RTs to sounds are modulated, during the exposure to audiovisual asynchrony, in the direction predicted by the "temporal shift" hypothesis. Alternate explanations for this result (Section 1.2.2.1) include realignment and Kalman filter-like updating (e.g., Burge, Ernst, & Banks, 2008) of estimate bias (see Section 4.4 & 5.1).

4.2 Methods

4.2.1 Observers

Nineteen naïve observers (14 female, mean age of 24 years), with normal hearing and normal or corrected-to-normal vision (by self-report) took part in this experiment. They received a 5 pound (U.K. Sterling) gift voucher in return for taking part in the experiment which was conducted in accordance with the Declaration of Helsinki.

4.2.2 Stimuli

In block 1 (synchronous), the visual stimulus (a 5 degree-wide green ring) was presented (on a Philips 107-E Monitor, 85 Hz) together with a 1000Hz beep (delivered, at 75.5 dB(A) via 2 loudspeaker cones, one located on either side of the screen) for 24 ms out of every 1000 ms. In block 2 (asynchronous), and for the VA group of observers, there was a 224-ms gap between the onset of the visual stimulus and the onset of the auditory stimulus. For the AV group, the onset of the auditory stimulus appeared 84 ms before that of the visual stimulus. Asymmetrical asynchronies were chosen on the basis of previous research demonstrating that, from an early age, people tolerate more asynchrony when the visual signal arrives before the auditory signal (Lewkowicz, 1996; McGrath & Summerfield, 1985). There is extensive evidence (e.g., see Dixon & Spitz, 1980; van Wassenhove, Grant, & Poeppel, 2007; Vatakis & Spence, 2012; Vroomen & Keetels, 2010) suggesting that humans tolerate more asynchrony when vision leads than when vision lags. This is probably due to the fact that situations arising in the natural environment involving audiovisual asynchronies, visual signals tend to appear before auditory signals (e.g., your friend calling to you from 100m away) because sounds are perceived first in physiological synchrony. The same parameters were used for the 8 re-exposure pairs of audiovisual stimuli between the sets of 4 RT trials. Background white noise was presented continuously at 62.5dB(A) during the entire experimental session.

4.2.3 Procedure

The experiment consisted of two blocks (synchronous and asynchronous). As it is still not clear how long the temporal adaptation ''aftereffect'' lasts, a blocked experimental design was used in which the synchrony condition (where no adaptation was expected to occur) always preceded the asynchrony condition. In this way, the chances of obtaining any undesirable carry-over effects were reduced. In block 1 (synchronous), observers were exposed to brief visual and auditory stimuli (both lasting 24ms) appearing in synchrony for 5 min. To indicate the change from exposure to test phase, the instruction “press” appeared on the screen. Next, 4 unimodal visual or auditory stimuli (that were identical to those used in the exposure phase) were presented pseudo-randomly with an inter-trial interval of 1500-3500ms. The observers, sitting 50cm from the screen, had to press the spacebar as rapidly as possible whenever they perceived either stimulus. Eight ‘re-exposure’ pairs of audiovisual stimuli appeared, following the instruction “do not press” between every 4 RT trials until a total of 112

RT trials (56 visual and 56 auditory) had been presented. In Block 2 (asynchronous), a brief silent gap was introduced between the visual and auditory stimuli during the initial exposure and the re-exposure stimuli. The two experimental blocks followed exactly the same structure. However, half of the observers (VA group; 9 observers) were exposed to audiovisual asynchrony, in the second block, where the auditory stimulus was always delayed by 224 ms with respect to the visual stimulus, while the other observers (AV group; 10 observers) were exposed to asynchronous stimuli where the auditory stimulus led by 84ms (see Figure 4.2 for further details). In order to ensure that the observers were attending to the stimuli, they had to detect occasional oddball stimuli (25% of the total) that were different from the "standard" (i.e., a thinner ring together with a 1300 Hz beep). Detection performance on this task was near-perfect. In order to both familiarize the observers with the task and reduce as much as possible the effects of practice on the RTs, a practice block preceded the main experimental blocks. This block consisted of 60 trials (30 to-detect-as-fast-as-possible acoustic and 30 visual stimuli) that were presented at unpredictable intervals.

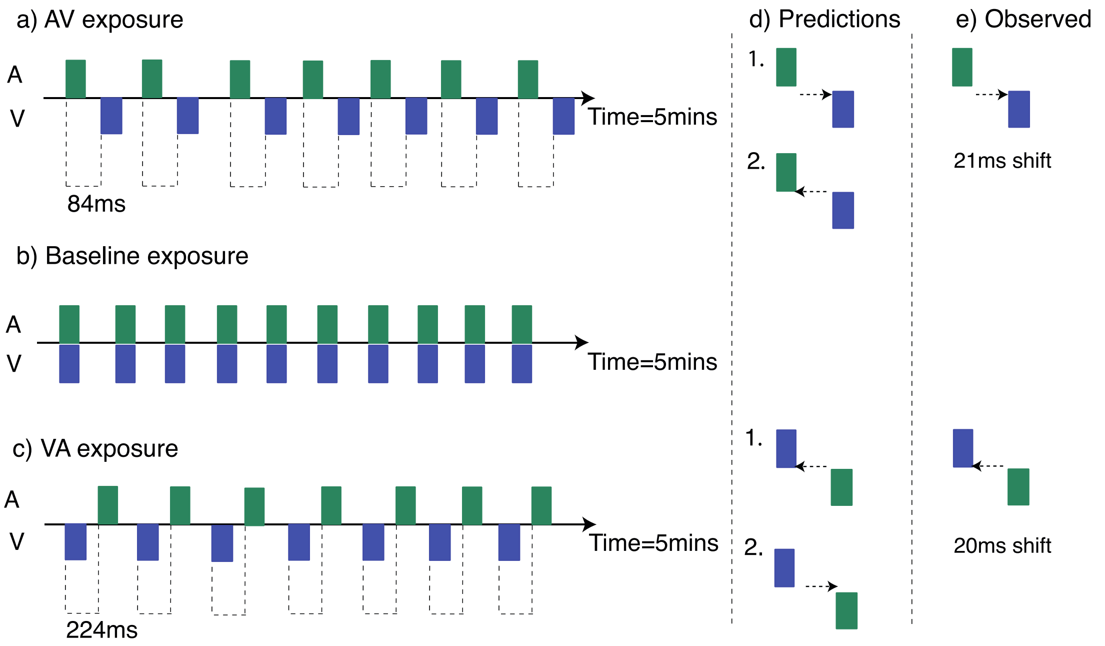


Figure 4.1. **Predicted and observed after-effects. The predicted and observed results are presented for the cases in which the audio signal leads (upper row) and lags (lower row) with respect to the visual signal. d) According to Prediction 1 (see the 2nd column), and following the observation that auditory stimuli – green – can be attracted toward visual stimuli – blue – in the temporal domain, specific modulations in auditory RTs are expected as a result of temporal recalibration. By contrast, RTs to visual stimuli (see Prediction 2 in the 2nd column) could be modulated by exposure to audiovisual asynchrony as a consequence of the fact that visual (but not auditory) arrival times, are modulated by the distance of an event from the observed. The various different predictions expressed in this panel should not, however, be treated as exclusive. e) As the observed change highlights, specific modulations of RTs were observed only in observers' responses to auditory stimuli.**

4.3 Results

In order to analyse whether adaptation to a specific audiovisual asynchrony influences observers' RTs to subsequent unimodal (visual or auditory) stimuli, observers' average unimodal (speeded) detection RTs (falling in the 100-450 ms range) after a 5-minute exposure to either synchrony (baseline) or asynchrony were calculated. An ANOVA, including two within-participants factors ('sensory modality': visual vs. auditory; and 'type of exposure': synchrony vs. asynchrony), and one between-participants factor (VA vs. AV group), revealed a significant three-way interaction [F(1, 17)=5.7, p=0.029]. This result suggests that the exposure to audiovisual asynchrony (VA or AV), but not to audiovisual synchrony, differentially affected observers' responses to the auditory and visual stimuli. Further analysis revealed that the two groups (VA and AV) differed in terms of the asynchrony effect (i.e., the result of subtracting the RTs in the synchrony condition from the RTs in the asynchrony condition), only in their speeded responses to sounds [t(17)=-4.2, p=0.0006], indicating that exposure to audiovisual asynchrony affected the speed with which observers responded to sounds, but not to visual stimuli [t(17)=-1.6, p=0.136].

Following the methodology used in previous studies (e.g., Fujisaki et al., 2004) "re-exposure" stimuli (that were all synchronous or asynchronous, depending on the block: synchrony or asynchrony, respectively) were alternated with the RT trials in the test phase (that is, after each synchrony and asynchrony exposure phase). This design made it possible to determine how the observers' RT changed over time as observers' exposure to a particular asynchrony increased. According to the Vincentization procedure (see Graighero, Fadiga, Umiltá, & Rizzolati, 1999; Ratcliff, 1979), each participant's RTs in the experiment were grouped into 6 "temporal bins" (or "periods") of approximately 18 trials each (Synch1, Synch2, Synch3, Asynch1, Asynch2, and Asynch3) and averaged for the VA and AV groups separately (see Figure 4.2). As for the previous analyses, all RTs falling outside the 100 - 450 ms range were excluded from these analyses. An ANOVA including the factors of modality (visual vs. auditory), time period (from Synch1 bin to Asynch3 bin) and group (VA vs. AV) revealed a significant main effect of time period [F(5, 85)=3.6, p=0.006], and significant interactions between time period and group [F(5, 85)=3.1, p=0.01], and between modality, time period, and group [F(5, 85)=2.6, p=0.03]. Additional analyses revealed that, while an effect of time period [F(5, 85)=3.1, p=0.01], and an interaction between time period and experimental group [F(5, 85)=4.6, p=0.001], were found when responses to auditory stimuli were considered, no such effects were observed on visual responses. This result reveals that participant responses to auditory, but not to visual stimuli were faster as a result of exposure to VA asynchrony and slower as a result exposure to AV asynchrony.

RTs from the last bin/period (that is, the last 36 trials; 18 visual and 18 auditory) in both the synchrony and the asynchrony blocks were taken from each modality and group in order to compare the "final" effect of exposure to synchrony with the "final" effect of exposure to asynchrony (see the arrows in Figure 4.2). An ANOVA, including the factors of sensory modality (visual vs. auditory), exposure (synchrony vs. asynchrony), and group (AV vs. VA), revealed a significant 3-way interaction [F(1, 17) 5.1, p=0.04], and a 2-way interaction between exposure and group [F(1, 17)=7.1, p=0.02], again indicating that exposure to audiovisual asynchrony modulated auditory RTs differently in the AV and VA groups. More detailed analyses confirmed this result: while an interaction between Exposure and Group was observed when considering auditory RTs [F(1, 17)=8.9, p=0.008], no trace of such an effect was observed when visual RTs were analysed [F(1, 17)=1.3, p=0.276]. Note that the effect reported fell in the range of the temporal shift effects previously found using other methods (e.g., Fujisaki et al., 2004; Vroomen et al., 2004). Finally, it should be noted that, when considering only the data from the three temporal bins in the synchrony block, there was a near-significant trend, in RTs, to decrease in the AV group (F(2, 18)=2.8, p=0.089; indicating that RTs were decreasing). This result may be considered to reinforce the notion that the responses in the AV group (the one showing, in the asynchrony condition, progressively larger RTs) were being modulated by exposure to asynchrony.

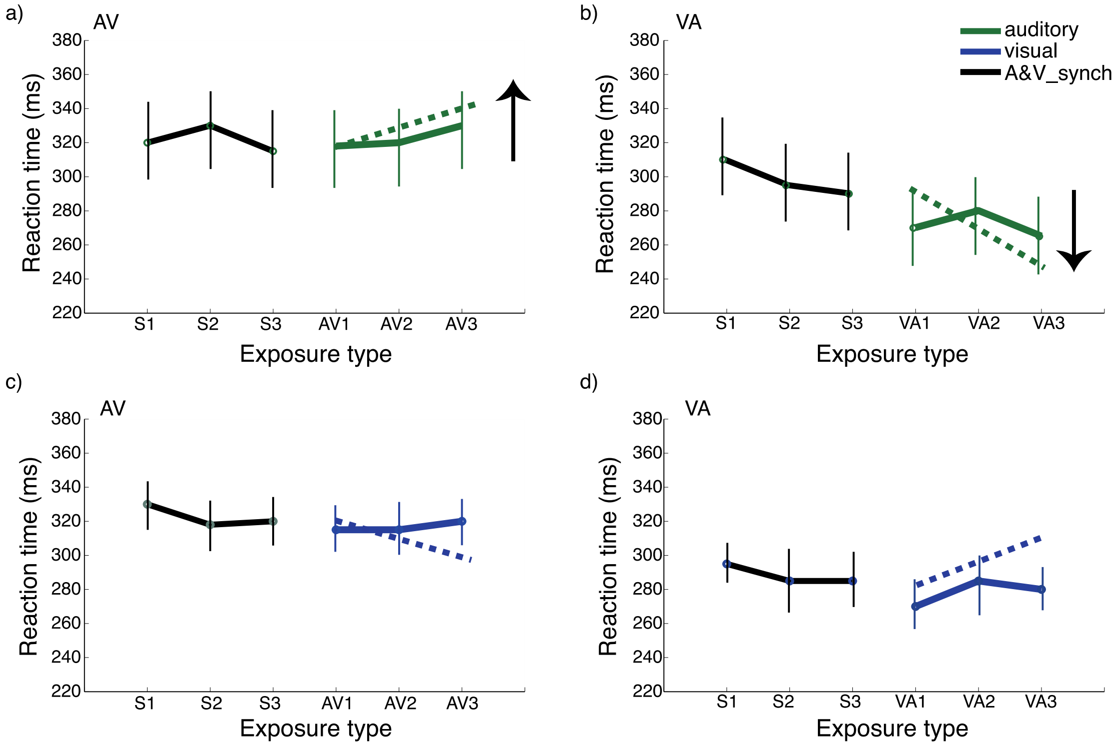


Figure 4.2. Responses over time. RTs (ms) are represented (on the Y-axis) for vision in figures c) & d) and audition a) & b), for each experimental group b) & d) for vision leading [VA] and a) & c) auditory leading [AV]. Error bars represent the standard error of the mean. The observers' responses over the course of the experiment were grouped into six different temporal bins – 3 taken following exposure to synchrony, in block 1, and the remaining 3 taken in block 2, following adaptation to audiovisual, VA or AV, asynchrony. Auditory RTs tended to be faster as the exposure (and re-exposure) to an auditory-lagging asynchrony increased and slower as the exposure (and re-exposure) to an auditory-leading asynchrony increased. No such pattern of results was observed for the visual RTs. The arrows indicate the difference between the last 18 trials in the synchrony (block 1) and asynchrony (block 2) conditions. This difference was significant when considering only the RTs to sounds.

The overall difference between the two groups was not significant for RTs to sounds in synchrony (t(17)=1.4, p=0.17), and nor was it significant for RTs to visual stimuli in synchrony (t(17)=1.7, p=0.10). This, and the fact that the effects of exposure were found exclusively in responses to sounds (not different in the prior synchrony condition), demonstrates that the effects of exposure to asynchrony reported were not due simply to a general (and unexpected) difference between the two groups.

4.4 Discussion

The most parsimonious explanation for the results of the experiment reported in the present chapter is that during exposure to asynchronous audiovisual stimulation, the speed at which the auditory signal is processed changes to enhance the subjective impression of audiovisual synchrony (see “Observed after-effect” in Figure 4.1). According to the data, this change is sufficient to modify the time at which observers are able to respond to sounds. Recent research using electrophysiological recordings has demonstrated, in line with the results reported here, that some specific event-related potentials (ERPs) involved in the detection of a sound (namely N1 and P2) can be observed earlier depending on the visual information that precedes them (Stekelenburg & Vroomen 2007; van Wassenhove, Grant, & Poeppel, 2005). Thus, the RT results reported here, when combined with the data from other tasks, such as those involving TOJs or SJs (see Cardoso-Leite et al., 2007), strongly suggest the existence of a mechanism that can re-adjust the temporal processing of incoming sensory signals that are related (e.g., by means of causality, as in the case of seeing the preparatory lip movement producing /b/ and subsequently hearing the related sound; see van Wassenhove et al., 2008) or tend to occur close in time (as in the current experiment). Importantly, similar temporal effects have also been reported during the perception of our own actions and their consequences (see Engbert, Wohlschläger, Thomas, & Haggard, 2007; Stetson et al., 2006), perhaps indicating that the human nervous system tends to bind (in time) the asynchronous events that are, for whatever reason, related (e.g., causality, see Winter, Harrar, Gozdik, & Harris, 2008, or temporal proximity). Speculatively, it could be that a 'prior' – specifying that vision provides more reliable information regarding the absolute time of occurrence of an event – is applied to all possible cases of audiovisual asynchrony (perhaps even including the case in which audition leads vision), thus inducing a certain modulation of the auditory processing.

The absence of any effect in visual RTs following exposure to temporally misaligned audiovisual signals, though, does not necessarily imply that the processing of visual stimuli is never altered. It may, for example, be that the effects of temporal adaptation to asynchrony influence the processing of auditory and visual signals in different ways, and/or perhaps at different stages of neural processing (e.g., early vs. late). Another alternative explanation for the absence of any effect in vision could be that the auditory processing was a more suitable candidate for a temporal shift because of its lower reliability (due to the presence of non-masking white noise) with respect to the visual processing. This hypothesis opens up the possibility of, in the future, studying how the reliability of the perceived stimuli may, or may not, influence the mechanisms underlying temporal recalibration (see Di Luca et al., 2009).

It should, however, be born in mind that the vision often provides a more precise estimate of when a distant audiovisual event took place (see Section 4.1). This would also explain perfectly well why it is precisely the processing of sound that appears to be shifted in time. It is worth highlighting that the effects found, even when being observed just in one of the two pre-exposed modalities, are likely to be the consequence of the exposure to a very specific asynchrony (RTs increasing or decreasing as the exposure to an AV or VA asynchrony, respectively, augments), and not only to “asynchrony” or some other factors, as it could perfectly happen in the only previous study using RTs in a temporal recalibration paradigm (Harrar & Harris, 2008). Whether or not performance will be optimal as a result of this temporal readjustment will depend on each task/situation. Changing the speed at which sounds are processed would lead to a more accurate percept of when a distant multisensory event occurred (i.e., closer to its real time of occurrence), but it would also make simple detection sub-optimal under certain conditions (e.g., when sound detection is slowed down as a result of exposure to AV asynchrony; see Figure 4.2).

There might also be other ways of changing the speed of information flow during the perception of asynchronous signals. Even if the current data point to the idea that temporal adaptation influences unimodal (i.e., auditory) processing at a putatively early stage (see also Fujisaki et al., 2004; Stekelenburg & Vroomen, 2007; van Wassenhove, et al., 2005), it could be that the exposure to a specific asynchrony also induces a shift in, for example, the criterion used to decide how much sensory evidence is required in order to ‘detect’ a sound. Consistent with this latter hypothesis, adaptation to, for example, VA asynchrony would allow the system to accept less evidence in order for a sound to be detected, thus moving the criterion for the detection of a sound (but not the neural transmission speed) to an earlier point in time. Further research will be needed to elucidate whether adaptation to audiovisual asynchrony implies a genuine change in the speed at which the sensory signal is transmitted through the early stages of information processing (in a similar manner as in, for example, van Wassenhove et al., 2005) and/or a more “abstract” mechanism, such as a change in the criterion for detection. However, what should be clear from the results reported in this chapter is that temporal recalibration induces temporal shifts that are unimodal in nature and can be measured in a task as simple as speeded detection.

There is an apparent contradiction between the results reported in this chapter and the literature on temporal ventriloquism, where the time at which an observer perceives a visual stimulus seems to shift toward the instant at which a sound appears (e.g., Bertelson & Aschersleben, 2003; Fendrich & Corballis, 2001; Morein-Zamir, et al., 2003; Vroomen & de Gelder, 2004). One possible explanation for this discrepancy may be that temporal ventriloquism does not really induce any temporal shift in the processing of the visual signal. The fact that temporal ventriloquism modulates the amplitude of visual evoked potentials such as P1 and N1 but not their latency can be seen as providing evidence against the "visual temporal shift" hypothesis (Stekelenburg & Vroomen, 2007). Further research is needed in order to gain a better understanding of the mechanisms underlying temporal ventriloquism (and adaptation). Both effects illustrate the flexibility of the nervous system to update and combine multisensory information. The results reported here clearly demonstrate that changes in simple detection latencies to sound (perhaps reflecting one of the most basic human responses to a stimulus from the outside world) are influenced by prolonged exposure (in the timeframe of minutes) to audiovisual asynchrony.

The fact that responses to the auditory signal were modulated by adaptation to asynchrony is in line with findings of Di Luca et al. (2009) demonstrating that when auditory signals are presented from loudspeakers that are not co-located with the visual signal (as in this chapter for which loudspeakers flanked the screen, not-co-located) the auditory signal is the one that adapts. Similar to the current results they found a change in the RT to auditory stimuli and no change to visual stimuli (Experiment 3 with non-co-located stimuli). Di Luca et al. used headphones, which introduced a spatial conflict (see Zampini, Shore, & Spence, 2003a). Here sound stimuli were, in fact, presented from loudspeakers placed away from the monitor (while white noise was played in the background).

As pointed out by Di Luca et al. (2009), the difference in the pattern of results with different sound-presentation conditions indicates that what recalibrates depends on whether the signals are recognized by the nervous system as originating from the same location (though see Keetels & Vroomen, 2007, for a different interpretation of this). The choice of what is recalibrated could be attributed to a shift in trust (similar to the concept of weight described in Chapter 7) given to the sensory estimate containing a bias. As discussed in Section 4.1, in temporal tasks, the auditory signal is generally taken to be more reliable, unless in this case the intensity of the visual signal and the spatial conflict meant that the auditory signal was recalibrated. Bias cannot be directly measured in the signal so looking at how recalibration builds up over the exposure phase might be one way of gauging which signal is more likely to shift online and therefore contain a bias.

Trust can also be based on prior knowledge. In many ways this concept is analogous to the one used for describing cue integration in Section 1.2: in the cue combination signals are weighted according to the inverse of their variance to obtain the most reliable bimodal estimate (e.g., Ernst & Banks, 2002; Landy, Maloney, Johnston, & Young, 1995), whereas in the recalibration framework discussed here the relative trust given to a sensory signal is inversely related to the likelihood of bias of a sensory estimate.

4.5 Conclusions

Over time, with continuous exposure to an asynchrony response latency is adjusted to temporally realign auditory and visual signals with the auditory system undergoing the recalibrating shift. In order to measure whether the auditory change observed was due to bias in the auditory system, the shape of the function describing the recalibration process for auditory and visual signals is investigated in Chapter 5.

Chapter 5: Interrupting exposure to asynchrony leads to the separation of the sensory inputs, not recalibration

5.1 Introduction

Audiovisual temporal recalibration – as described in Chapter 4 – results from 'prolonged' exposure to a discrepancy in redundant audiovisual signals. Over time, the nervous system uses the redundancy in auditory and visual inputs and adapts to converge on a synchronous-state estimate of the signals. To achieve perceived synchrony, it deals with random errors that result from noise in visual and/or auditory timing and response estimates. But primarily recalibration uses the systematic error – that is, the change in the mapping between the visual and auditory signals provided by the artificial conflict – as used in the exposure phase of Experiment 4.1. In the natural environment such systematic error could result from a large distance between the observer and audiovisual event, providing auditory and visual signals that reach the observer asynchronously: the auditory arriving slightly later than the visual given the physics of the speed of sound and light (Fain, 2003; King, 2005; Spence & Squire, 2003). For such small asynchronies the nervous system assumes that the visual and auditory estimates provide noisy measures of some single distal event. Despite distance and other forms of discrepancy introduced in information processing, humans maintain accurate estimates of when something occurred. There is, however, no widely accepted model or explanation of how this recalibration process occurs.

The nervous system will recalibrate the discrepant inputs when sufficient evidence has been accumulated to suggest a common cause for the redundant but discrepant signals. Given this, it could be that continuous exposure is necessary to override the discrepancy as specifying two events. In other words, if the two discrepant signals occur together intermittently, do the statistics of the current environment promote a single cause for the redundant discrepant information, or not? Alternatively do they support a hypothesis of two distinct sources of information, apparently co-occurring?

A previous attempt to understand how recalibration occurs for a sensorimotor conflict proved successful in describing the rate of the adaptation process (Burge, Ernst, & Banks, 2008). By assuming that the nervous system operates as an optimal observer – i.e., it attempts to reduce the noise in its estimates of a dynamically changing world – Burge et al., applied a Kalman filter (see Welch & Bishop, 1997) to the problem of combining discrepant sensorimotor inputs in a reaching task. They were able to formally describe how adding noise, through feedback, to our already noisy human sensors, can modulate the course of adaptation that is observed.

In order to approximate the recalibration process – during exposure – and to understand what type of function can describe the process itself, the recalibration process was divided into segments. Each segment was followed by a test period. This sampling of unisensory RTs was an attempt to estimate the state of the sensory system across time. After each test phase the exposure recommenced. The intermittent sampling provided a means of estimating the shape, speed of – and signal to undergo – recalibration.

But how would the nervous system deal with prolonged but intermittent discrepant information? Everyday changes in our sensory-sensory mappings are probably best characterized by an accumulation of small changes rather than by large step-like changes (Baddeley, Ingram, & Miall, 2003; Kagerer, Contreras-Vidal, & Stelmach, 1997; Miyazaki, Yamamoto, Uchida, & Kitazawa, 2006). This creates a fundamental problem: to determine the appropriate response to a given error, the nervous system should respond differently depending on how much discrepancy was due to systematic error and how much was due to random error. The system cannot determine those relative proportions from any single observation, so a dynamic updating calibration method is required that balances the competing needs to average out random error while remaining sensitive to changes in systematic error.

5.1.1 Capturing the recalibration process during exposure

Burge et al., (2008) measured a particular aspect of sensorimotor recalibration. They did not assess which of the two systems recalibrated—the visual or the motor—but they confirmed another property of the Kalman filter, namely that the variance of the signals is a determining factor for the rate of recalibration. That is, although the variance of the signals should not determine the amount of recalibration in each signal, it determines the rate at which recalibration occurs.

Temporal recalibration provides an example of a recalibration process in which there is no direct access to the error contained in the dynamic estimate. As a result of this, in order to estimate – and predict – the nervous system's error at each moment in the updating processes, the approach requires more assumptions than spatial recalibration. One such assumption can come from prior information. If a sensory estimate had provided accurate (i.e., unbiased) information with a high probability in the past, it is likely to continue to provide unbiased information and thus may qualify for the stable, non-adapting component in the response to those estimates. Furthermore sensory motor recalibration provides direct feedback to the nervous system about the mapping between the visual and haptic (motor) components of the event. For sensory-sensory pairings, the interaction with the audiovisual event is passive. Observers do not interact with the object and yet the nervous system is recalibrated, eliminating the asynchrony. What is meant by passive? Take sensorimotor pairings which, on the other hand, allow feedback from interaction – i.e., catching a moving object will provide information on the temporal accuracy (bias) of the perceptual estimates involved. If there is bias, you will not catch the object. This can, in turn, be used to update the bias inherently associated with those signals. Burge, Girshick, and Banks (2010) suggest each estimate should be calibrated by an amount proportional to the estimate weight (as described in Chapter 1.1.2). For the audiovisual temporal recalibration explored in the experiment reported in this chapter, a relative weight can be estimated when comparing the responses to AV exposure to AV with noise.

In summary, two factors are examined here in order to determine if temporal recalibration can occur when it is necessary to interrupt the adaptation process, in order to obtain a dynamic representation of the adaptation of vision and audition.

The effects of temporal recalibration on response latency, as observed in the experiment reported in Chapter 4, demonstrate the timing of the entire auditory system (in this instance) is readjusted in the updating process (e.g., Fujisaki et al., 2004; Hanson et al., 2008b; Harrar & Harris, 2008; Miyazaki, et al., 2006; Navarra et al., 2007; Vatakis et al., 2007; Vroomen et al., 2004). Thus, the shift observed through recalibration can be described as a process of realignment, not remapping (see Di Luca et al., 2009). The work reported in the present chapter takes a novel approach in order to investigate what happens in real time in the process of temporal recalibration. A problem with an online measure is that the recalibration process is weakened after the sensory system has been deprived of exposure for some time (see Keetels & Vroomen, 2010; Machulla, Di Luca, Froehlich, & Ernst, 2012). The effect of recalibration on RT has been demonstrated a number of times (Di Luca et al., 2009; Navarra et al., 2009) and is clearly depicted in the results of Chapter 4 (see Figure 4.2). Based on this, the predictions for the current experiment are that adapting to an audio leading [AV] discrepancy should result in responses to the auditory input slowing down, while RTs to visual input should be speeded up. The response characteristics to [VA] exposure should be reversed, i.e., visual processing should slow down and auditory responses speed up, therefore resolving the temporal discrepancy.

Using a unisensory measure of recalibration makes it possible to assess what recalibrates and how rapidly recalibration actually occurs. In the experiments reported in the present chapter, simple RTs to auditory and visual unimodal stimuli will be used, to investigate whether this change in response latency can be observed when the exposure phase is interrupted.

Furthermore, the motivation underlying the research reported in this chapter was to see whether it is possible to observe recalibration when the exposure phase is interrupted, and use the intermittent sampling of RTs to measure which signal changes, and what type of function can describe that process of change. A second aim of the current chapter was to understand what happens to adaptation under different forms of asynchrony – constant or with noise. Changing the amount of noise in the asynchrony should allow the speed of adaptation for each signal to be assessed and potentially described by a Kalman filter model. If random noise in the adapting signals changes the rate of adaptation, then the current results can also be included in the optimal adaptation category described by a Kalman filter (see Burge, et al., 2008; Greenwald, Knill, & Saunders, 2005; Welch & Bishop, 1997).

5.2 Methods

The general paradigm described here is similar to that utilized in Chapter 4. However, fundamental differences in the exposure and measurement phases, and the apparatus used, require further elaboration.

5.2.1 Observers

Ten observers (four females) took part in the current experiment. All reported normal or corrected-to-normal vision and normal hearing. The average age of the observers was 23.9 years. The observers were naïve as to the purpose of the experiment and received a compensation of 8 Euros per hour for their participation.

5.2.2 Apparatus and stimuli

The stimuli were generated using Matlab (Mathworks) and were presented via a custom built device. A multi-channel sound card (M-audio 1010LT) was used together with identical power amplifiers to drive both LED and sound-attenuating headphones to ensure sub-millisecond precision and accuracy. The audio card produced 20-ms sinusoids with frequencies of 2,000 Hz and 150 Hz, respectively, for the audio and visual stimuli, with 5ms linearly ramped onset and offset. Stimulus intensities were 76 dB SPL for the auditory signals and 41 cd/m2, for the visual. The red LED was 5mm in diameter. Auditory stimuli were presented using noise-attenuating headphones (KOSS QZ99). A loudspeaker positioned away from the apparatus produced continuous white noise (74 dB SPL). In order to ensure that subjects paid attention throughout the presentation of the stimuli, 1–9 oddball stimuli with a deviating sound frequency (1,000 Hz) or light intensity (76.5 cd/m2) were presented among the standard stimuli during the exposure phase.

5.2.3 Design

The experiment was run in 3 sessions for the different types of exposure: visual leading [VA] by 150ms, auditory leading [AV] by 150ms, or auditory leading with the asynchrony selected from a random distribution μ=150ms, σ=5ms [AV_noise]. All observers ran all three conditions. Noise was added to only one direction of asynchrony to ensure that the experiment remained a reasonable length. The order of the different exposure phases was counterbalanced across observers.

Each session consisted of 20 experimental blocks. Each block of trials in turn consisted of an exposure phase and a measurement phase (see Figure 4.1a). In the exposure phase for the first five blocks, observers were exposed to an asynchrony of 0ms irrespective of the type of exposure. In the subsequent 15 blocks, observers were exposed to an asynchrony. This could either be with vision leading [VA] or with audition leading [AV] or [AV_noise]. Finally, in the measurement phase, observers made speeded responses to unimodal auditory and visual stimuli.

5.2.4 Procedure

Exposure phase: Each exposure block in the exposure phase lasted considerably less time than in the studies described in Chapters 4 and 6 (exposure period in the latter two of 5 minutes), although the entire phase lasted approximately 20 minutes. The experimental procedure in the current chapter is depicted schematically in Figure 5.1. During each exposure phase, the observers were presented 5 initial blocks with an asynchrony Δ=0ms, to ensure that the audiovisual signals were perceived as originating from one object/event (see Figure 5.2a-c). In the subsequent 10 blocks observers were presented with audiovisual stimulus pairs with an asynchrony Δ=150ms (Δ=150ms + noise for [AV+noise]) either [AV] or [VA][7]. In the final blocks an asynchrony step change occurred again returning to Δ=0ms. Between each stimulus pair, a random pause (between 250 and 400ms offset to onset) was inserted, in order to render the cross-correlation between the visual and auditory series of events unique. To ensure that observers attended to both the auditory and visual stimuli during the exposure phase of the experiment, they were asked to count the number of oddball

---

[7] In the first block of asynchronous exposure, the asynchrony changed from 0ms to 150ms in steps of 50ms. This is customary for the temporal recalibration paradigm (Fujisaki, et al., 2004; Navarra, Hartcher-O'Brien et al., 2009; Heron et al., 2008b).

stimulus pairs (between 3 and 8 brighter lights or lower pitched sounds were presented in each exposure phase).

Test phase: The test phase, directly following exposure, consisted of 10 trials of auditory and visual stimuli, 5 presented to each modality at random. The observers had to respond as quickly as possible by pressing a response button when they saw or heard a stimulus. No feedback was provided. Once the response was given, there was an inter-trial interval of 700 ms for 10 trials. After these 10 trials the next exposure block began. In total a session lasted approximately 40 minutes.

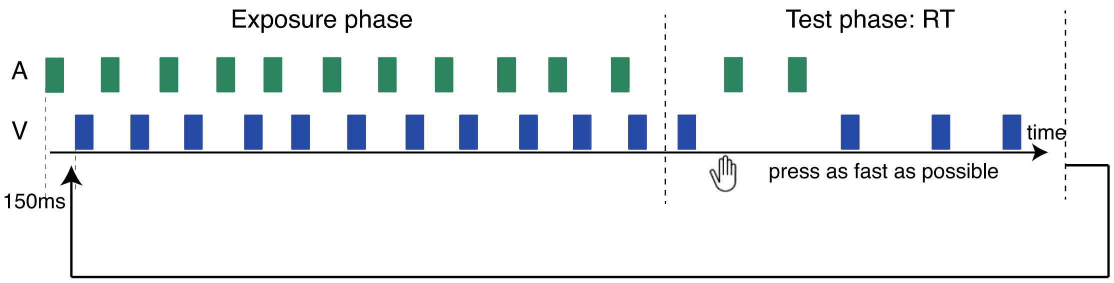


Figure 5.1. Depiction of experimental paradigm and stimulus presentation for a single block (where exposure was to an audio leading asynchrony of 150 [AV] of the experiments in this chapter. The exposure phase was interrupted by the RT measurement to unisensory auditory and visual stimuli. The arrow indicates that each response time collection was followed by the re-exposure for 20 blocks in total.

5.2.5 Data analysis

RTs slower than 900ms, faster than 100ms, or more than 3 standard deviations from the participant's average, were excluded from the analysis. All data were analysed for auditory and visual RTs separately. The average RT across observers was taken as the mean of the log transform of the RTs (as described by Carpenter & Williams, 1995). To smooth the RT data across the measurement blocks, a running-window averaging procedure was implemented for visualization purposes.

5.3 Results

Responses for the oddball procedure used to verify that the observers attended to both sound and light in the exposure phase were highly accurate (98% mean correct responses overall).

5.3.1 RT over time – does recalibration occur?

The first question addressed in the present chapter was whether recalibration of response latency actually occurs when the exposure phase is interrupted for sampling purposes. The observers were exposed to both directions of asynchrony, audio leading [AV] and visual leading [VA], with an additional manipulation of a random noise jitter added to the audio leading condition. The data for the [AV] condition will be discussed first.

The RTs – sampled during exposure to [AV] asynchrony – were transformed such that the noisy non-linear RT distribution approximated a log-normal distribution (see Carpenter & Williams, 1995, for further elaboration). This has implications for the analyses such that the response latencies are now assumed to come from a normal distribution. The data were then averaged across observers.

The smoothed response latencies during [AV] adaptation for audition and vision are presented in Figure 5.2a). The effect of adaptation or recalibration can normally be approximated by an exponential response function (Wark, Lundstrom, & Fairhall, 2007).

The prediction for adaptation, as measured in sensorimotor adaptation (see Burge et al., 2008), is data that can be approximated by a Kalman filter. However, given the data obtained, this is not the case. Instead, the data were fit with a gain function as described below. The RTs from block 6 to 15 for the asynchrony recalibration phase – i.e., adaptation to asynchrony – and those from baseline and post (i.e. blocks 1-5 and 16-20 respectively) were smoothed using a running window-averaging procedure with a window size of 3 data points either side of the point being smoothed. The smoothed data were then fitted with a gain function to approximate the adaptation phase for vision and audition separately. Figure 5.2 a) demonstrates the decrease in auditory response latencies over time, and the concurrent increase in visual responses (blues) across the session. The observed trends for both sensory modalities occur in the opposite-than-predicted direction if recalibration had taken place: auditory responses should slow down in order to bridge the audiovisual asynchrony while visual responses should speed up. Furthermore, the adaptation process is extremely noisy despite normalizing the RT distributions (although not noisier than those reported in Chapter 4), as seen in the non-exponential, oscillatory trajectories for both vision (blue) and audition (green).

If the nervous system realigns visual and auditory inputs in order to eliminate the temporal discrepancy, adapting to an AV asynchrony would require that responses to auditory signals decrease and those to visual signals increase. In the current chapter the exposure phase presenting audiovisual asynchrony was interrupted intermittently for measurement of RTs. However, interrupting the exposure to AV asynchrony negated this effect. The change from block 1-5 represents the baseline adaptation to 0ms asynchrony. From blocks 6-16 the exposure asynchrony was presented and then a step change was introduced again from 16-20. There is no significant change for either vision or audition in the adaptation asynchrony blocks (6-15). In the [AV] exposure condition, the discrepancy presented was constant and therefore the nervous system should not have had a problem in accounting for the discrepancy. Interrupting the exposure phase inhibits the realignment process. The dotted vertical lines represent the step changes in asynchrony from 0ms to 150ms in the exposure part of the experiment.

When random noise jitter is added to the adapting asynchrony, the response trajectories for both modalities change at a faster rate than in the no noise [AV] adaptation condition (see Figure 5.2b). Noise in the adapting asynchrony therefore determines the rate at which recalibration occurs. It should be noted, however, that although in this condition the auditory response changes in the direction predicted for recalibration, the visual response latencies change in the same direction thus suggesting that noise also adds to the noisiness of the estimate.

5.3.2 RT cumulative distribution affected by noise

To gauge the overall effect of noise in exposure on the response latencies, the cumulative distribution of the RTs was taken and examined as a function of mean RT for the three adapting asynchronies, [AV], [AV_noise], and [VA]. These cumulative distributions provide insight into the rate and range of change as a function of the noise in the adaptation phase. In Figure 5.2d the auditory cumulative RTs show a greater change when noise is added to the [AV] condition, as opposed to the condition with no noise in both the [AV] and [VA] adaptation conditions.

In Figure 5.2d the effect of noise on the response latencies to vision and audition can be observed as a steeper change in the cumulative response for the [AV_noise] condition than in either of the non-noise conditions [AV or VA]. For visual RTs, on the other hand, as noise in the adaptor increased from [AV] to [AV_noise] the slope of the cumulative response function became shallower.

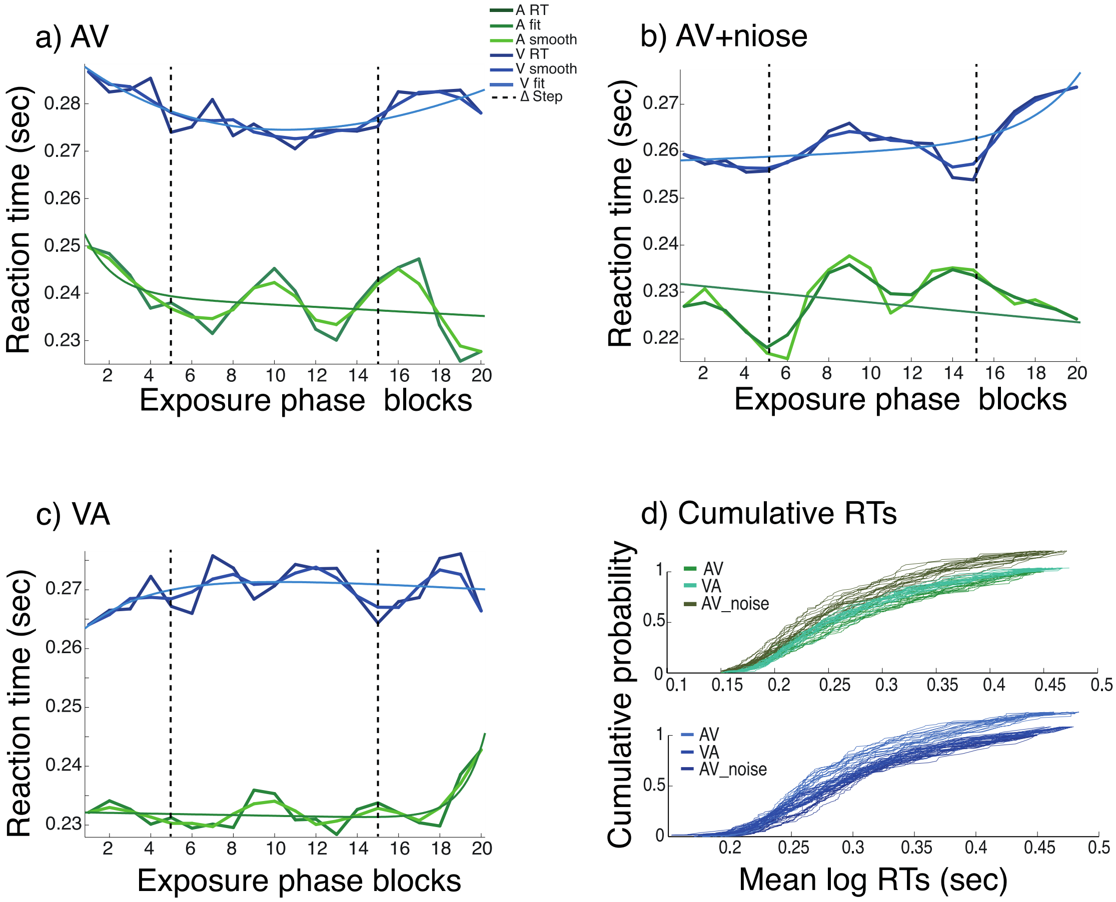


Figure 5.2. RT data in seconds from Experiment 5.1 for auditory (dark green) and visual (dark blue) responses is a) given as a function of recalibration phase blocks for audio leading [AV} asynchrony. The smoothed data deviate more strongly for the auditory (lighter green) than for visual (lighter blue) response latencies. The best-fit describing the change in RTs is a gain function f(x) = a*exp(b*x) + c*exp(d*x). b) RT data in seconds, smoothed and fit with a gain function are shown after AV asynchrony with a random with random noise μ= 150ms, σ=5ms [AV_noise}. The change in RT as a function of block is greater in this condition. c) The RT data, smoothed and fit with the gain function are presented following exposure to visual leading asynchrony [VA]. The dotted vertical lines represent the step changes in asynchrony (Δ). d) The cumulative distribution of RTs as a function of mean RT to auditory (greens) and visual (blues) signals in the test phase of [AV], [VA] and [AV_noise] exposure. The change observed for the different exposure types is colour coded for the two modalities. The cumulative distributions for all 10 observers are given.

The RTs to [VA] adaptation show no change across the adaptation period for vision or audition Figure 5.2 c). The smoothed data and the fits are represented in exactly the same manner as those in Figure 5.2a and b.

The Kalman filter model of recursive dynamic adaptation can best account for the change in variance in the estimates over time. In order to explore whether this aspect of the current data were able to demonstrate Kalman like changes the change in variance for the AV condition and AV plus random perturbation condition for the two step changes in asynchrony are presented in Figure 5.3.

As seen in Figure 5.3, the variance of RT responses changes slightly over time. For the initial five blocks the observers were exposed to an asynchrony of 0ms. On the 5th block the asynchrony increased from 0-150ms. It is evident for the AV condition (solid line) that at this step, and the second step back to 0ms, the variance changes distinctly for the auditory stimulus, mimicking the change in discrepancy. At step one it increases and then decreases on the re-introduction to the 0ms delay.

The change in the visual responses is more sluggish – perhaps mimicking the more sluggish change in impulse response function for this system. However, what can be deduced from the current variance data is that the variance profile for auditory and visual responses are equal at the end of the process. In the AV+noise condition there is no change in the visual variance for the entire 0ms exposure phase. When the discrepancy is increased to 150ms the variance fluctuates strongly until the end of the experimental period, irrespective of step. This is similar to the auditory variance response pattern: the steps do not cause any jumps in the variance response.

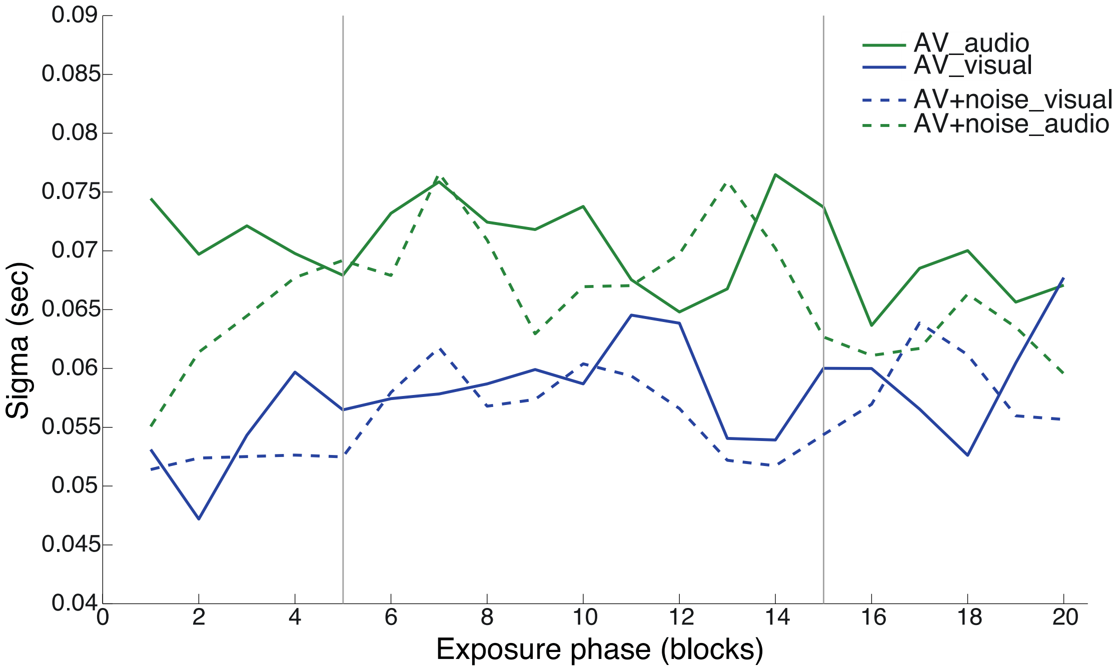


Figure 5.3. The RT variance (σ^2) from Experiment 5.1 is plotted as a function of exposure block for adaptation to AV (solid lines) asynchrony and AV+noise (dotted lines). Auditory σ (green) and visual σ (blue) represent the variance across trials and observers. The step changes in the discrepancy between auditory and visual stimuli (exposure asynchrony) occurred at block 5 and block 15, given by the grey vertical lines.

For adaptation to VA asynchrony, there was no change in the response latencies to vision or audition over time. In other words, no recalibration occurred in this condition. RTs to auditory and visual signals do not recalibrate in the predicted direction, if the exposure phase is interrupted by intermittent test phases. In light of the lack of recalibration observed, applying a Kalman filter is more hypothetical than practical. However, the fact that the data are modulated by systematic noise may suggest that a different method for measuring recalibration may bring the process to light.

5.4 Discussion

5.4.1 Summary of findings

Contrary to the predictions, interrupting the exposure to asynchrony appears to have forced the nervous system to treat the two inputs as arising from separate events. This conclusion is supported by the fact that the small change in RTs over exposure blocks occurred in the opposite direction than required for temporal realignment of the signals, i.e., establishing a stronger two event cause. Adding noise to the exposure asynchrony had a systematic influence on the range of response latencies for the auditory RTs. The cumulative distribution of responses highlight a steeper change compared to the no-noise AV adaptation condition. This is interesting as it further suggests that the additional noise strengthened the system's decision to treat the auditory and visual input as arising from two events, more quickly.

This attempt to estimate the shape of the function describing temporal recalibration provided invaluable insight into the fact that the exposure to conflict needs to be continuous if the nervous system is going to integrate and realign the timing of things to come. What dictates whether recalibration *occurs* appears to be whether the events arise from the same source or not (e.g., Bedford, 1989; Meredith & Stein, 1993; Vatakis, Ghazanfar, & Spence, 2008). The effect null-recalibration effect observed is supported by results of Machulla, Di Luca and Ernst (2010), which demonstrate that without top-up, the recalibration effect is weaker than that reported in other studies. Furthermore, Machulla et al. (2010) examined the decay of recalibration and found that the waning of the effect occurs in the absence of stimulation. Perhaps the lack of recalibration in this chapter, resulting from an interruption to the exposure phase, provides another perspective on this decay phenomenon. Taken together, these results suggest that the system needs to establish a correlation between what is seen and heard, and maintain this representation before it will recalibrate to an audiovisual asynchrony.

One of the motivations for experiment 5.1, i.e., whether audiovisual recalibration can be described by a Kalman filter-like process was unable to be addressed by the results. Given the fact that no adaptation occurred under the interrupted exposure phase conditions, the Kalman filter was not applied as a best fit of the change in response latencies.

5.4.2 Kalman filter and response latency – not an optimal measure

One issue with attempting to implement a Kalman filter on RT data is that, even when the RT distribution has been normalized using a log-transform, the assumptions of normality implicit in the Gaussian noise based model of the Kalman filter, cannot be generalized to the findings presented in this chapter. Moreover, estimating the random noise element from RT data is fraught with an additional inflation of noise, due to the non-normal distribution from which the RTs are drawn.

5.4.3 Can we access the bias and ascertain what is recalibrated?

Humans adapt to discrepancies in the mapping between sensory inputs in order to maintain a coherent and robust representation of the environment. For sensory-sensory temporal recalibration, providing feedback about the discrepancy or error is not so simple. Two main factors influence the rate of this recalibration in a sensory motor context: the extent to which the current mapping is reliable (mapping uncertainty) and the extent to which the visual feedback is reliable (feedback uncertainty) (Burge et al., 2008). In terms of our optimal observer asked to resolve a static audiovisual temporal conflict (see Chapter 7), it was possible to model the combined perceptual estimate using MLE. For dynamic processes such as adaptation, the Kalman filter has been shown to be a useful model for describing sensorimotor recalibration (Burge et al., 2008).

The results of Experiment 5.1 manifest a critical calculation for the nervous system: whether visual and auditory stimuli are derived from one environmental source or more. Without being able to establish a correlation over the exposure phase and infer whether the current stimulation is due to one or more environmental sources, there would be no point to cue integration. The calculation of source likelihood must work despite the variable neural and physical transmission times that occur with visual and auditory stimulation and processing. Although the results do not approximate the function describing recalibration they shed light upon the properties of the mechanisms that determine temporal coincidence of visual and auditory stimuli. The successful operation of these mechanisms may facilitate many important forms of integration, such as allowing us to attribute spoken words to the appropriate person.

5.5 Recommendations for future research

The Kalman filter approach would have extended the Bayesian model of the optimal observer (see Chapters 1 and 7) to dynamic estimates. The attempt to apply such a model to temporal recalibration was one way of exploring the effect at level of the nervous system and its state estimation. Perhaps the more useful level of description for the recalibration process is the neural level. One indirect way of examining how recalibration is affected by signals arising from a common cause is to measure the transfer of recalibration to a non-adapted signal. This is the question that will be investigated in the following chapter.

5.6 Conclusions

The attempt to estimate the shape of the function describing temporal recalibration provided invaluable insight into the fact that the exposure to redundant yet temporally conflicting stimuli needs to be continuous if the nervous system is going to integrate the audio and visual components and realign the timing of things to come. In order to probe whether recalibration only occurs for signals deemed to arise from the same event the following chapter explores the transfer of recalibration for simultaneity estimates to a change in auditory frequencies from the adapted to the test stimuli.

Chapter 6: No transfer of recalibration to an un-adapted frequency tone

6.1 Introduction

Previous findings, including those reported in Chapter 4, have established that temporal recalibration leads to a shift in the perceived timing of asynchronous audiovisual stimuli such that, a subsequent pair of stimuli appear more synchronous than upon previous presentation (e.g., Di Luca et al., 2009; Fujisaki et al., 2004; Hanson, Heron, & Whitaker, 2008a; Harrar & Harris, 2008; Heron, Whitaker, McGraw, & Horoshenkov, 2007; Keetels & Vroomen, 2007; Navarra, Garcia-Morera, & Spence, 2012), however, it is still contentious as to whether recalibration occurs when the multisensory components could be inferred to originate from distinct events (see Keetels & Vroomen, 2007; Roseboom & Arnold, 2011; Yarrow, Rosenboom, & Arnold, 2011). One study exploring the influence of spatial separation of stimuli (an important cue indicating that stimuli arose from the same event – see Chapter 1.1) demonstrated that recalibration occurs independent of other such cues imply that the multiple sensory signals arose from the same event (Keetels & Vroomen, 2007). Findings from a recent study that explored the influence of potential complementary visual features, such as spatial frequency, demonstrated an influence of such cues (Heron, Roach, Hanson, McGraw, & Whitaker, 2012). Heron et al. (2012) also investigated the role of contextual cues. No recalibration was observed when the test stimulus contained a change in stimulus feature (in their case, pitch and spatial frequency). These effects provide support for a low-level recalibrational mechanism that operates on spatiotopic coordinates for the visual component of the event. Evidence of spatiotopic selectivity in the recalibration mechanism suggests that recalibration across different tonotopic bands in the auditory system may well be band specific (see Figure 6.2).

The results reported in Chapter 5 indicate that interrupting exposure to the discrepancy changes the nervous system's response strategy from one of recalibration (see Chapter 4) to one of stimuli separation as a result of exposure. The implications of the findings reported in Chapter 5 are that intermittent exposure to an asynchrony actually forces the nervous system to infer that the two signals arose from two separate events. If the nervous system is unable to establish that the signals in question arise from the same distal event, then realigning them in order to achieve perceptual synchrony could be detrimental to the observer (see Ernst & Di Luca, 2011). A nervous system that correctly identifies the reason for the discrepancy, i.e., a systematic error which should be corrected or a random error that should be ignored, increases the chance of maintaining event and object coherence despite the extrinsically and intrinsically generated delays with which the nervous system is continually faced.

In the broader scheme of things, all species face the imperative for identifying, recognizing, and localizing the sensory signals essential for their survival and reproduction. Failure to do so will obviously ultimately endanger their survival. Thus to ensure accurate estimation and representation of events and object, species have evolved to use certain cues contained in a signal in order to be able to define its identity (and location). A few examples of such cues include the characteristic – resonant – frequency of an object, the temporal pattern in speech and binaural disparities for sound localization used by humans and especially various species of owl (Knudsen, Knudsen, & Esterly, 1984; McAlpine, Jiang, & Palmer, 2001; Moiseff, 1988). These different physical attributes make the processing of large amounts of information more efficient, with a higher probability of the correct meaning being assigned to the cues.

To explore whether recalibration depends upon signals being inferred by the system to arise from a common event, the current experiment exploited the role that fundamental frequency plays in aiding the nervous system to make this object assignment. Evidence of the role auditory frequency plays in object identification and recognition comes from a number of studies, both animal and human (see Carr & Konishi, 1990; Kayser, Petkov, Lippert, & Logothetis, 2007; Moore, 2003; Nityanander & Bee, 2011). According to Carr and Konishi, signal frequency plays a fundamental role in recognition and identification. They suggest that to highlight the importance of this, the auditory system has evolved to preserve this cue throughout the auditory pathway from cochlear to cortex (Cynx & Shapiro, 1986). On the other hand, temporal features of events – i.e., the rate at which something flickers, flutters, or changes, the order in which two things occur, the duration of events – these stimulus features change dynamically and continuously in the natural environment. Flexible processing of these stimulus properties is therefore equally essential for any organism’s survival.

The nervous system is often capable of simultaneously processing the timing of events and correctly parsing available cues, such as frequency in order to maintain event or object coherence (see Roseboom & Arnold, 2011). The hypothesis posed for the current experiment is that since frequency represents a stable cue, while temporal properties have a high degree of uncertainty associated with them, recalibration should occur for single frequencies but not transfer to a non-adapted frequency. Furthermore, the continuous exposure provided during the adaptation phase allows the nervous system to establish a correlation or prediction that the two signals originate from a single event (Di Luca et al., 2009). In Chapters 4 and 5, choice RTs were used to assess the influence of recalibration on the individual sensory systems. In the current chapter, a more direct measure of the nervous system's differential latency between audition and vision was introduced – simultaneity estimates. Implementing a second measure for temporal recalibration provides a more thorough and detailed picture of the process of recalibration itself.

Predictions for recalibration of the adapted and non-adapted frequency signals, using an SJ task, can be seen in Figure 6.2. These predictions are based on the assumption of a dedicated asynchrony mechanism that operates early in the auditory pathway (see Heron et al., 2012). However, it could also be the case that for recalibration, the auditory system does not segregate the events based on a frequency identifier (see Figure 6.2 b), but generalizes for the auditory system as a whole. The two auditory frequencies were selected in order to test for tonotopic adaptation. That is, in order to activate different frequency region responses of the auditory system, maximizing the chance of the two signals being assigned to different events. Further support for the prediction that recalibration should not transfer is provided by the fact that a fundamental organizing principle of the auditory system is frequency selectivity or tonotopy (e.g., Talavage, Sereno, Melcher, Ledden, Rosen, & Dale, 2004).

The experiment to be reported in this chapter was designed to test whether recalibration is specific to the adapted frequency pair, or transfers to a stimulus change that occurs within-modality but should activate distinct frequency bands or tonotopic regions of the auditory system. Temporal recalibration in this instance would therefore mean that the non-adapted frequency would adapt also, unless the coding of temporal properties maintains the frequency segregation applied at all levels of auditory processing, from cochlear to auditory cortex. In order to assess the identity dependent influences on the nervous systems ability to flexibly update the temporal relationship between two signals, a more direct measure of the differential latency of the nervous system for vision and audition was used – a simultaneity judgment (see Yarrow & Arnold, 2011).

6.2 Methods

6.2.1 Observers

Twelve naïve observers (8 female), with a mean age of 22 years (SD=4.5), with normal hearing and normal or corrected-to-normal vision took part in this experiment. They received a 5 pound (U.K. Sterling) gift voucher in return for taking part in the experiment which was conducted in accordance with the Declaration of Helsinki 1964.

6.2.2 Stimuli

Exposure phase: The visual stimulus was a green annulus (outer ring 5° and an inner ring of 2.25°) and was presented on a Philips 107-E Monitor, 85 Hz. It appeared for

23.4 ms which was the equivalent of 1 frame. The adaptor auditory stimulus consisted of a 3,500Hz sinusoidal signal (delivered, at 75.5 dB(A) via 2 loudspeakers, one located either side of the screen). The duration of the auditory stimulus was equal to that of the visual stimulus. During exposure, the audiovisual stimulus pair was repeated every 1,000ms for a total duration of 30,000 ms. In the asynchronous exposure phase for the VA condition, there was a 224-ms lag between the onset of the visual stimulus and the subsequent onset of the auditory stimulus. For the AV condition, the onset of the auditory stimulus appeared 84 ms before the onset of the visual stimulus, producing the adaptation lag. Asymmetrical asynchronies were chosen on the basis of previous research showing that, people tolerate more asynchrony when the visual signal arrives before the auditory signal (van Wassenhove et al., 2005).

In the test phase, a second non-adapted auditory frequency was also paired randomly with the visual annulus. This auditory test stimulus was a 1,000 Hz sinusoidal signal of equal duration to the auditory adaptation stimulus. The inter-trial interval was randomly selected from a distribution between 1,500 and 3,500 ms. White noise was presented continuously in the background at 62.5 dB(A) during the entire experimental session (i.e., exposure and test phases).

6.2.3 Procedure

The experiment consisted of three blocks defined by the type of exposure in the exposure phase (synchronous, VA asynchronous, and AV asynchronous). In each block, the exposure phase was followed by a test phase in which observers performed a SJ, where they indicated whether the audiovisual pair was simultaneous or not.

6.2.3.1 Exposure phase

The exposure phase was identical to that described in Chapter 4 (see Section 4.2). However, one auditory signal with a frequency of 3,500 Hz (represented in green in Figures 6.1 & 6.2) was coupled with the visual stimulus. In the test phase this stimulus was again used as the adapted auditory frequency.

6.2.3.2 Test Phase

At the end of the 5-minute exposure phase, there was a 2 sec pause followed by 10 seconds of top-up adaptation. Following this re-exposure, observers performed a SJ task in which they were required to indicate whether the audiovisual stimulus pair was perceived as simultaneous or not. Before each test pair, the circle fixation marker changed to a cross, to indicate to the participant that the test pair would follow. After re-exposure, an audiovisual pair was presented with a SOA between the two signals randomly chosen from 9 values between -280 and +280 ms [-280 -180 -100 -40 0 40 100 180 280]. Each session consisted of 72 test trials. The trials were divided into six blocks. A block contained one repetition of the 12 different test SOAs presented in a random order. The adaptation phase frequency stimulus is also used in the test phase and SJ results from this frequency are indicated throughout this chapter in green. SJs resulting from the non-adapted auditory frequency signal coupling are given in yellow.

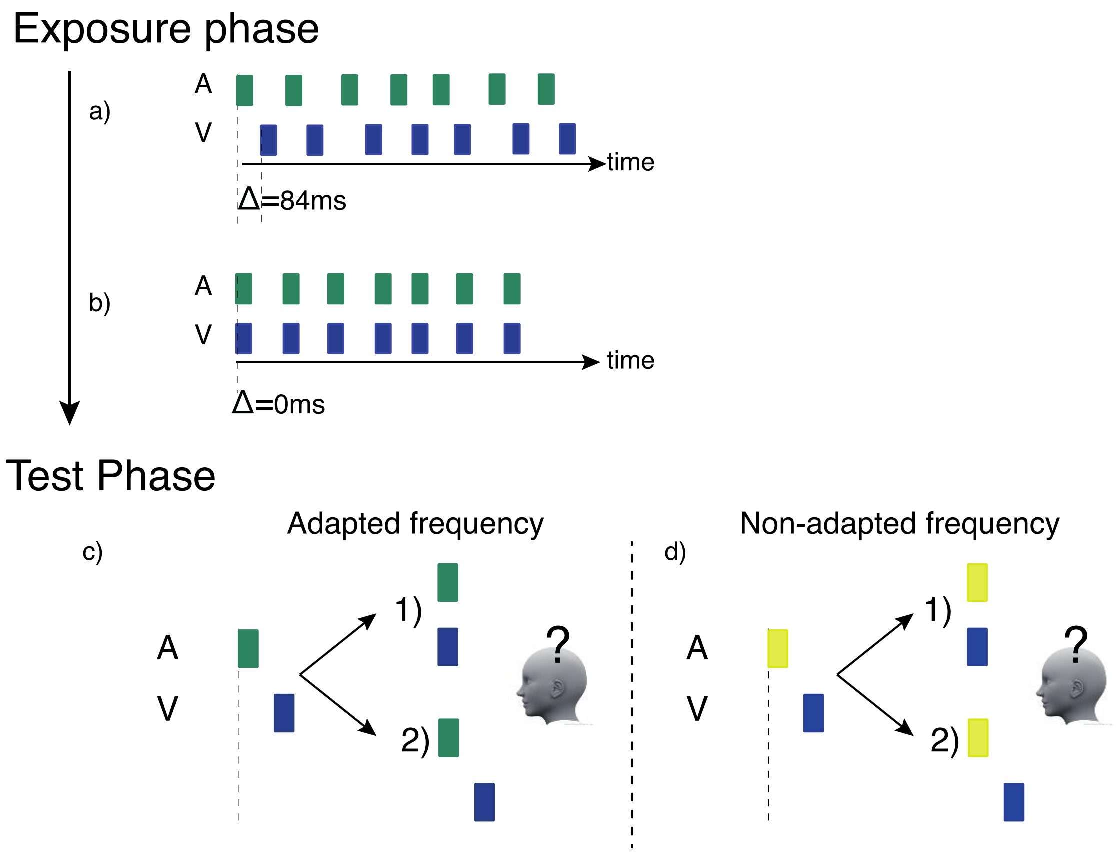


Figure 6.1. A schematic representation of the experimental design for both exposure and test phases in Experiment 6.1. The audiovisual stimuli pairs in the exposure phase: a) Adapting to an [AV] audio-leading asynchrony of 84ms, b) adapting to [S] synchronous audiovisual signals. c) The test phase following the exposure period in which observers perform the SJ task for single audiovisual pairs. The audio component of the test pair consisted of the adapted auditory frequency or e) a non-adapted frequency auditory signal.

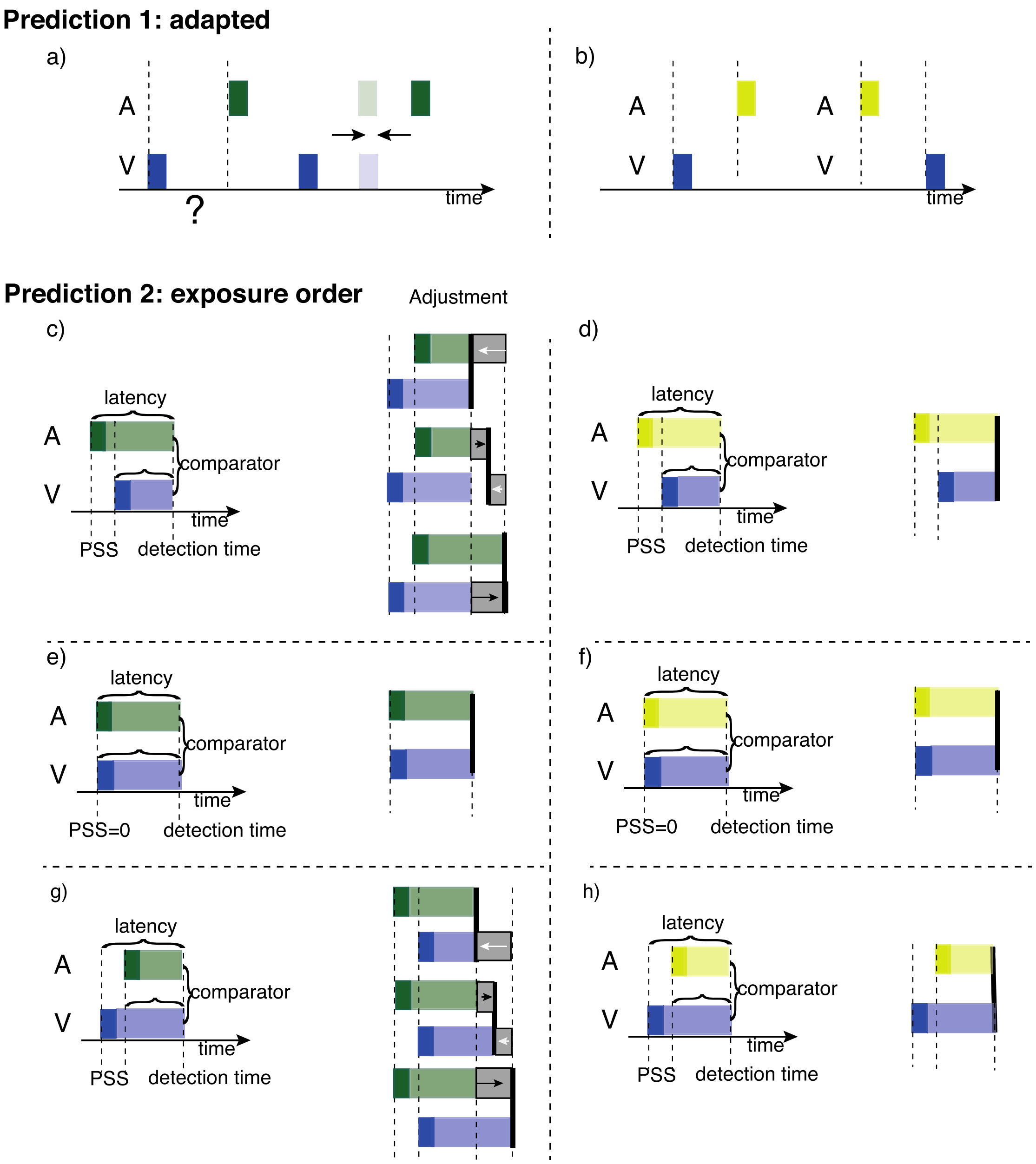


Figure 6.2. Predictions for a) the adapted – green – and b) the non-adapted frequency – yellow – audio signals. Visual signals are schematically represented in blue. The predicted shifts required to realign the signals in time changes for the different order of exposure and is given for c) audio leading [AV] in the case of the adapted frequency and d) for the non-adapted frequency signal. To align the signals perceptually, a delay could be imposed on the auditory system to resolve the asynchrony or the visual system could speed up information processing. e) The control condition exposure to [S] synchrony provides the baseline to disentangle whether the change in PSS is due to fatigue in the neurons or a temporal shift. f) The effect for test pairs with the non-adapted frequency signal after [S] adaptation. g) For the visual leading [VA] condition the visual signal could be processed with a delay compensating for the comparator discrepancy or the audio system could increase its speed of response and align the two signals. h) For the non-adapted frequency no change is predicted.

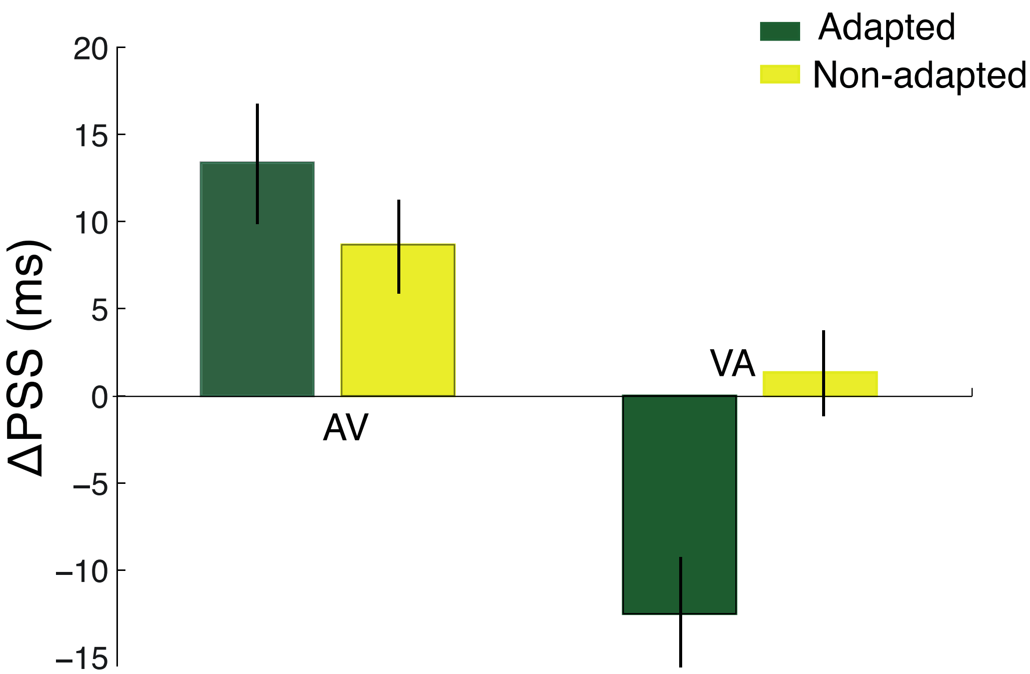
6.2.4 Data analysis

Psychometric functions were calculated for each participant and condition by fitting a normal Gaussian function to the 'proportion simultaneous' responses. The point of subjective simultaneity (PSS), which indicates the SOA at which the observers were maximally uncertain about whether the two stimuli were simultaneous or not, was also calculated from each function as the mean of the distribution. The discrimination threshold or the just noticeable difference (JND) was computed from the reciprocal of the steepness of the best-fitting Gaussian function. For the simultaneity estimates, one JND equates to one standard deviation of the Gaussian distribution underlying the psychometric function. The fit parameter of interest here, however, is the PSS, as adaptation should produce an effect on perceived simultaneity of the stimuli indicated by the PSS.

6.3 Results

6.3.1 Effect of Adaptation

The initial prediction for the simultaneity estimates was that the adapted frequency trials would show a shift in the PSS when normalized for adaptation to [S] simultaneous audiovisual stimulus pairs. Moreover, the PSS for the non-adapted frequency estimates were not expected to shift relative to the baseline [S] PSS. All analyses were performed on the normalized PSS estimates, i.e., Δ PSS. ΔPSS was calculated as $\text{PSS}_{AV/VA} - \text{PSS}_{S}$. Figure 6.4 shows the change in ΔPSS ms for the adapted (green) and non-adapted (yellow) frequency trials for the normalized AV and VA conditions.

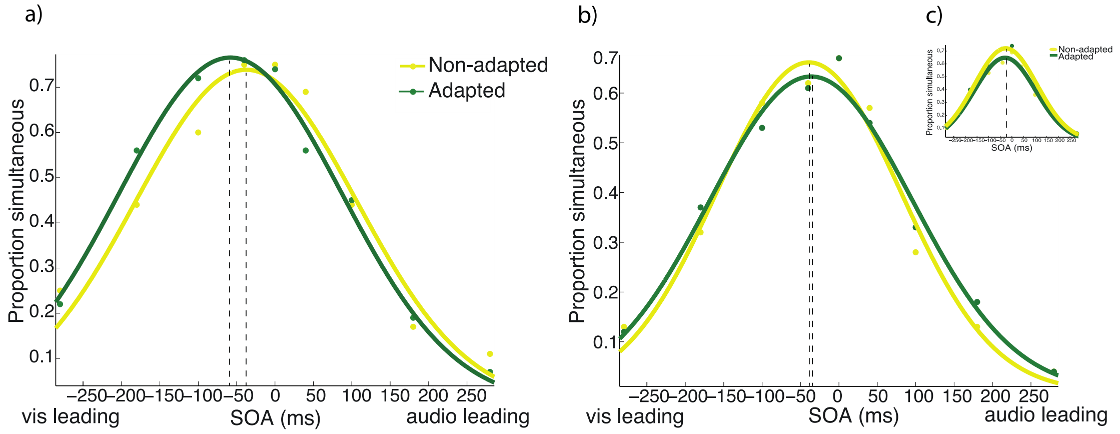


<u>Figure 6.3</u>. The normalised PSS values ΔPSS (ms) for the adapted (green) and non-adapted (yellow) frequency trials from Experiment 6.1. The left column represents the ΔPSS (ms) for adaptation to AV-S and the right column represents the ΔPSS (ms) after adaptation to VA-S.

A 2 x 2 repeated measures ANOVA comparing frequency (adapted vs. non-adapted) and order (ΔAV vs. ΔVA), provided evidence for a significant main effect of adaptation (adapted or non-adapted frequency) – F(1, 11)=7.99, *p*=0.01) and, moreover, a significant interaction between adaptation and order (ΔAV vs. ΔVA) – F(1, 11)=6.12, *p*=0.03. To understand the nature of the interaction, Bonferroni post-hoc comparisons were performed. These revealed a significant difference in the order condition VA for adapted vs. non-adapted frequency estimates t(11)= –2.76, *p*=0.01. It is clear from visual inspection of Figure 6.4 that the change in ΔPSS (ms) is significantly greater in the adapted direction for the adapted (green) versus non-adapted (yellow) frequency estimates (right column) in the ΔVA condition.

6.3.2 Effect of order of exposure and adaptation

The non-normalized PSS (ms) for the adapted and non-adapted frequency signals is derived from the Gaussian fits to the proportion simultaneous data. In Figure 6.4, the Gaussian fits to the proportion simultaneous responses as a function of SOA, are presented for exposure to a) audio leading [AV], b) visual leading [VA] and synchronous presentation of the AV pairs provided in the insert graph. For the difference in order of stimuli in the exposure phase: [VA] vs. [AV] no difference was observed for the audio leading condition. The maximal shift between adapted (green) and non-adapted (yellow) frequency estimates is demonstrated a) in the [VA] condition. That is, for the VA pair, all observers' PSS values differed significantly in the predicted direction (mean ΔPSS_{VA} ± SE=-10.26 ± 7 ms for the adapted frequency estimates; paired samples t-test comparing adapted and non-adapted frequency estimates was significant: $t(11)=4.55$, $p=0.0019$).



<u>Figure 6.4</u>. The normal Gaussian fits to the proportion simultaneous responses for adapted (green) and non-adapted frequency (yellow) estimates from Experiment 6.1: a) following exposure to [AV] audio leading asynchrony; b) to visual leading [VA] asynchrony. The fit for the proportion simultaneous responses c) following synchronous exposure (inset) for adapted and non-adapted signals is given in the inset graph. The dotted lines represent the PSS or the mean of the normal distribution.

In order to perceive the auditory and visual stimuli as simultaneous, the auditory stimulus had to be presented 26 ms earlier after audio-leading exposure than after visual-leading exposure. The magnitude of this effect, that is, a difference in PSS corresponding to 8.6% of the 224 ms asynchrony during exposure, is in line with other reports concerning the recalibration of audiovisual simultaneity (Fujisaki et al., 2004, 12.5%; Keetels and Vroomen, 2007, 6.5%; Vroomen et al., 2004, 6.7%). In both the [AV] and [S] exposure phases the fits are clearly no different, as seen in Figure 6.4 B) and inset.

6.4 Discussion

6.4.1 Frequency-selective recalibration

Temporal recalibration to an audiovisual asynchronous event is preserved for the adapted frequency of the auditory component. Exposing observers to short term adaptation produces a significant frequency specific shift in the PSS for visual leading asynchrony. However, the current study applied uni-directional adaption – adapting to the (3,500Hz) frequency auditory component, and then testing AV stimulus pairs containing both this frequency and the non-adapted frequency signal (1,000Hz). A further conditional on the change in PSS is that all conditions had an initial PSS of approximately -40ms – which may represent the acknowledged difference in audiovisual processing that should occur for physically synchronous events (see Fain, 2003). Irrespective of this, an effect of adapted frequency was only present in the visual leading condition.

The first finding suggests that humans have an adaptation mechanism that is auditory frequency selective, which first tonotopically maps, and then updates, the relationship between the two signals. This is in agreement with recent findings (Heron et al., 2012) indicating that recalibration is stimulus feature specific for features such as contrast (although see Navarra et al., 2012). The data reported in this chapter also shed light on when and why the nervous system decides for recalibration as opposed to segregation. Previous research exploring other aspects of auditory temporal processing (Carr & Konisha, 1990; Moore, 2003), suggests that temporal properties are coded tonotopically, thus providing support for the tonotopic-selective shift observed in this chapter. In additional support, Moore (2003) examined neural firing rate in non-human primates for adaptation to individual auditory frequencies. He demonstrated that frequency-specific changes in the neural response occur for the adapted frequency but not for other frequency signals: after normal adaptation to specific frequencies, the targeted neurons adjust their temporal responses. Given the frequency specific responses of the auditory system it is plausible that a mechanism of readjustment would also work on tonotopic gradients for the auditory signals in temporal recalibration. The findings reported here, coupled with those reported in Chapter 4, suggest that temporal recalibration occurs by a readjustment of the auditory system's timing for frequency–specific information. As suggested in Section 6.1.2, the auditory system uses tonotopic distinctions for object recognition and identification. One way to maintain identity while updating the temporal properties of the event is to recalibrate in an auditory frequency specific manner. Given (auditory) identity is coded in frequency terms, it is a biologically and evolutionarily plausible response.

The current findings implicate order of stimuli in the exposure phase as important. This order effect observed could be the result of two factors, one more methodological and the other drawing on an evolutionary or learned response mechanism. The first account offered here for the asymmetry in recalibration ([VA] adapting more than [AV]) is that the exposure asynchrony was larger in the vision leading condition. The asynchrony values chosen were selected because they signify the average temporal window of integration for the different orders of stimulus presentation (see Recanzone, 1998; van Wassenhove et al., 2005; Vroomen et al., 2004). This may, however, account for the magnitude difference in the effects observed. The size of the temporal recalibration effect is proportional to the adapting asynchrony – on average 6-12% of the adaptor (Fujisaki et al., 2004; Keetels & Vroomen, 2007; Vroomen et al., 2004). This more methodological explanation is, however, only one possibility. The second possibility is that the nervous system has evolved to deal with visual leading asynchronies (van Wassenhove et al., 2005; Vatakis & Spence, 2007) over an observer's lifetime. This would support the fact that the nervous system is more likely to bind and recalibrate stimuli when vision leads, i.e., the [VA] exposure condition. Furthermore, a visual leading audiovisual asynchrony has implications for events and actions in the real world, such as speech. In the case of speech, for example, the onset of the lip movement precedes the voiced auditory signal. An individual's voice is identifiable by its fundamental frequency. Thus the lack of frequency transfer may also represent a strategy used by the nervous system to identify the incoming signals.

6.4.2 Generalizability and limitations of the results

The findings presented in this chapter provide support for a frequency-selective dedicated asynchrony mechanism that updates temporal properties of events early in visual and auditory sensory pathways (see also Heron et al., 2012). One way to probe this claim further could be to expose observers to a broadband auditory signal (i.e., white noise), coupled with a visual annulus and then in the test phase present auditory signals constituting the individual frequency components. If recalibration for these signals is observed then recalibration is tonotopic, not necessarily object based. However, given that other researchers have also reported that frequency selective neural responses is preserved after adaptation (see Carr & Konishi, 1999; Heron et al., 2012; Moore, 2003), it is possible to speculate that updating temporal properties of events occurs while frequency information is preserved by the nervous system.

Whatever the mechanism responsible for this preservation of frequency, it is interesting to explore the functionality of the effect. As such, it must be noted that the results could also be interpreted in terms of a novel versus familiar stimulus effect – one adaptor frequency and the familiar adaptor and novel non-adapted auditory frequency. Stimulus novelty irrespective of frequency could also account for the effect observed in this chapter. In relation to the claim that auditory identity is coded via frequency for natural events, it is interesting to compare the current findings – using simple pure tone stimuli, to the results of recalibration studies that have used speech (Roseboom & Arnold, 2011; Vatakis, Navarra, Soto-Faraco, & Spence, 2007). The study conducted by Vatakis et al. (2007) demonstrated that recalibration resulted in a change in JND, effectively broadening the window for multisensory integration. Roseboom and Arnold also used speech stimuli but explored the effect of identity. Their findings suggest that complex audiovisual speech signals produce identity specific recalibration. Observers' perceived timing of events was changed only for the adapted visual speech stimuli. These findings support the current claim for recalibration of identity-coded signals.

Furthermore, the frequency-specific effect observed in the experiment reported in this chapter is in line with findings that both temporal order and duration adaptation occurs spatiotopically, i.e., for events coded in real world coordinates, and is not generalized to stimuli in non-adapted spatiotopic locations (Burr et al., 2007; Heron et al., 2012, though see also Johnston et al., 2006). Therefore, simultaneity is not the only temporal property to demonstrate such coordinate specific adaptation as discussed in Section 6.1.2. Duration adaptation arguably occurs spatiotopically suggesting that the external object rather than the retinotopic location is important to object coherence (Burr et al., 2007). In a perceptual motor loop, the key function of the auditory system is object identification based on frequency and object location based on interaural time and level differences. Exploring what is affected by recalibration and using frequency selectivity assumes the mechanism available to the auditory system in temporal tasks is still tonotopic.

Previous research exploring spatial recalibration demonstrated a tonotopic selective remapping of space for audiovisual recalibration (Recanzone, 1998). These findings from spatial recalibration, coupled with the results reported in the present chapter provide evidence for auditory system segregation being maintained, both for temporal and spatial properties of the environment. The extent to which recalibration (either spatial or temporal) maintains tonotopic coding is unequivocal.

Whether temporal recalibration generalizes depends on the environmental properties tested (Di Luca et al., 2009; Fujisaki et al., 2004; Heron et al., 2012). In their seminal study published in *Nature Neuroscience*, Fujisaki et al. (2004) showed that temporal recalibration transfers across different visual stimulus features, such as colour and pattern. Coupled with the work of Di Luca et al. (2009), it is clear that some transfer occurs for a recalibrated nervous system. However, tonotopic selectivity remains un-ameliorated. Although Navarra et al., 2012 have demonstrated a transfer across frequencies, this is probably due to the fact that the sensory system to undergo the recalibrational shift in their work was the visual – i.e. transfers across auditory frequencies, and in this experiment the auditory stimulus adapted (see Di Luca et al., 2009).

6.4.3 Future work

The study presented in this chapter demonstrates an auditory frequency specific recalibration to a temporal discrepancy between audiovisual signals. Open questions remain as to whether this effect occurs (i.e., low-level, frequency specific or beyond auditory cortex – generalises across frequency) for exposure to two superimposed frequency-band signals then demonstrating recalibration when tested on either alone. To really understand fully what is happening with frequency specific adaptation, there are many manipulations which could tease apart the exact nature of this effect.

6.5 Conclusions

The results of Experiment 6.1 help us to understand how invariant representations are formed and maintained during the process of recalibration. Features or cues, such as stimulus frequency, are important for coding object identity, and representation of position. They are therefore preserved as segregated during recalibration.

6.6 Mechanisms of temporal recalibration: A general discussion

Aftereffects in perceived simultaneity following exposure to asynchronous stimuli have been related to the functioning of different types of mechanisms. In general, an aftereffect could either produce an adjustment of the processing of the individual stimuli or it could be obtained with a mechanism specific to the multisensory pair of stimuli - that is which operates via a remapping between the two signals themselves. Notice that the changes operated by the two mechanisms can operate either through adaptation or recalibration, as we will see below. To re-establish temporal coherence (recalibration), a unisensory adjustment would have to change the timing of stimuli by modifying the latency at which the stimuli become available for perception and action. The latency of the leading stimulus in the pair presented during exposure should increase and/or the perceptual latency of the second stimulus should decrease. Adaptation could operate for adjustment mechanisms, too. For example, repeated exposure to a single stimulus could change the sensitivity to the stimulus and thus change its perceived timing. This would also have an influence on perceived simultaneity. Instead, for the remapping mechanism, the unisensory processing of stimuli is not affected, but there is a change in the relationship between the stimuli. This strategy has been described in terms of recalibration, where a temporal window of simultaneity can vary depending on several attributes of the multisensory stimulation (Kopinska & Harris, 2004). Another way of looking at this possibility is with a series of coincidence detectors that undergo adaptation. These multisensory neurons shift their temporal alignment preference during the exposure period to the discrepant stimuli. Roach et al. (2010) proposed a neural population code where relative timing is represented by the activity of a number of neurons. Each neuron is tuned to a single asynchrony. Consistent with the effect of adaptation, repeated exposure to a constant asynchrony reduces neuronal response of the tuned neurons. Such a mechanism has been studied in the nucleus laminaris of the barn owl (analogous to the mammalian medial superior olive) where neuronal delay lines and coincidence detection, as in the Jeffress model are used to code interaural time differences due to the off-axis location of sound sources (Jeffress, 1948; Knudsen & Konisha, 1978).

The two mechanisms—adjustment and remapping—make distinct predictions regarding the extent to which the aftereffects of recalibration can be measured using single stimuli and can transfer to other pairs of stimuli. If aftereffects in simultaneity perception are achieved by remapping there should not be any registered difference with unisensory stimuli; this is because the remapping is specific to the relative timing of the stimuli in the exposed modality pair. On the other hand, if recalibration is achieved with a unisensory adjustment, an effect should be observed whenever the signal undergoing adjustment is involved, even if it is paired with a new stimulus. The transfer of recalibration to other types of stimulus pairs (which were not presented during exposure) would indicate that a mechanism based on unisensory adjustment is involved in generating the aftereffect. Results are somewhat varied, but the evidence in favor of a remapping mechanism is weaker and can be accounted for on the basis of newer findings. Harrar and Harris (2008) found no aftereffects in modality pairs other than the one exposed, nor changes in unisensory properties. However, both Navarra et al. (2009) and Di Luca et al. (2009) reported that the perceived onset times of auditory signals are either speeded up or slowed down (relative to their visual counterparts) following exposure. Similarly, Di Luca et al. found a change in RT and an effect on perceived simultaneity for non-exposed modality pairs. Not only does perceived timing of auditory and visual stimuli change after exposure to an asynchrony, but an aftereffect is also observed for visuotactile and audiotactile pairings (see also Heron et al., 2008). These findings suggest that exposure to multisensory asynchrony may influence the peripheral processing of information. Similar mechanisms have been proposed to explain audiovisual speech perception (van Wassenhove, Grant, & Poeppel, 2005) and the 'sound-induced illusory double flash' phenomenon (Shams, Kamitani, Thompson, & Shimojo, 2001). Here, however these changes in sensory processing are long lasting, as it has been shown to happen in several cases of perceptual learning (Bao, Yang, Rios, He, & Engel, 2010; Yang & Maunsell, 2004). The transfer of the aftereffect to other stimulus pairs that were not presented during exposure and the influence of the context in eliciting an aftereffect indicate that a neuronal mechanism based on top-down adjustment of sensory processing is likely involved.

If there is an adjustment of the sensory processing for one or both stimuli the question naturally arises as to what the factor is that determines the modality or stimulus type that undergoes recalibration. There are several feasible scenarios:

- It is possible that when resolving temporal conflicts the nervous system possesses a mechanism that recalibrates only one of the sensory modality. While the results of the experiments reported in Chapters 4 and 5 demonstrate a shift for only a single auditory frequency, other researchers have measured changes in the motor component (Sugano et al., 2010). Di Luca et al. (2009) show instead that the timing of both auditory and visual precepts can potentially be adjusted.

- The recalibrated modality is the one carrying less precise temporal information. For audiovisual stimulus pairs, it is a common belief that the auditory system codes temporal information more precisely (e.g., Recanzone, 2003). In this case, vision should normally be the modality affected by recalibration, as audition has higher temporal resolution than vision. Harrar and Harris (2008) found, instead, that RT to light was increased after exposure to lights-first audiovisual stimuli. The data of Di Luca et al. (2009) also shows that audition is not the only modality affected by exposure, as the presentation of sound with headphones does not change temporal precision of stimuli, but it influences which modality is recalibrated.
- If, instead, the nervous system could correctly identify the modality for which there is more variability in transmission delay, then audition would be the modality mostly affected by recalibration, touch should be intermediate, and vision should be considered as a temporal reference. Accordingly, Chapter 3 demonstrated that RT to auditory stimuli changed after exposure. Results from the study by Di Luca et al. (2009) add to this result showing that the modality undergoing recalibration could vary depending on the experimental setting, thus the nervous system must rely on some other parameters to decide which stimulus to recalibrate.

- The mechanism that changes perceptual latency could be functionally constrained to only decrease (or increase) the perceptual latency of stimuli. In this case the adjusted stimulus should be consistently the leading one of the asynchronous pair during exposure in case of an increase of latency, or the second one in the case of a decrease of latency. Result are mixed. Data from Di Luca et al. (2009) show only decreases of perceptual latency for the second stimuli presented during exposure. By contrast, the results reported in Chapter 4 and Navarra et al.'s (2009) results instead show both an increase and a decrease in RT to auditory stimuli after asynchronous exposure when compared to synchronous exposure. The functional limitation of the recalibration mechanism cannot therefore be the only factor that determines which of the two stimuli to recalibrate.
- The last possibility is that the stimulus or the modality undergoing recalibration is not fixed but determined each time depending on the situation. If the goal of recalibration is to establish accurate sensory estimates, then the choice of which modality to recalibrate should not be arbitrary but follow a precise rule. In the case of temporal estimates, accuracy means that the processing latency is known, so the signal that is less accurate should be recalibrated more. The accuracy of a temporal estimate, however, cannot be judged directly as the signal itself does not provide information about the processing latency. Accuracy can only be estimated on the basis of previous experience with the signals. If a sensory estimate had provided accurate information with a high probability in the past (i.e., unbiased), it is likely to continue to provide unbiased information and thus can be "trusted". Trust can

> only be expressed as a prior belief, as only feedback can be used to build up or update such a prior belief. For an ideal observer, the amount by which each estimate should be calibrated is determined by a weighting scheme where the weights are inversely proportional to the belief that the estimates are accurate.

Psychophysical data in this regard is still scarce, thus strong conclusions cannot be supported yet. One way of looking at the problem from another point of view is with neural recordings. The SC is one of the nervous system structures involved in multisensory processing and recording of activation after asynchronous exposure offers some indications of how perceptual changes are obtained (Yu et al., 2009). Recordings before and after repeated exposure to asynchronous audiovisual stimuli indicate a change in the neural responses to the two stimuli so to make them overlap in time despite their asynchrony. The response to the second stimulus has a shorter latency and the response to the first stimulus has longer duration and overall higher response magnitude. These changes arise gradually as the stimulus pair is repeatedly presented and they last beyond exposure when stimuli are presented individually. It is yet to understand how these changes in neural activity map to perception.

Producing an integrated neural response from multisensory stimuli depends on the location and relative timing of the stimuli. Stimuli are treated as independent events if their asynchrony falls outside of the window of integration. If integration crucially depended on an exact match in time, most events that take place near the observer would not be integrated due to the difference in processing latency across the channels. Probably to avoid such temporal selectivity, neural recordings have shown that the width of the window exceeds hundreds of milliseconds into the second range

(Meredith, Nemitz, & Stein, 1987). Nonetheless, within the wide temporal window there exists an optimum asynchrony for which the multisensory enhancement on neural response to the stimulation is maximal.

In order to understand multisensory integration of redundant cues for other temporal properties, i.e. duration, Chapter 7 applies a more formal, mathematical approach to teasing apart the mechanisms underlying multisensory integration of redundant temporal cues.

Chapter 7: On the duration of uncertain times

7.1 Introduction

Imagine you are having a conversation and your partner pauses, momentarily. You try to estimate if they are simply taking a breath or if it is a subtle cue, signalling that it is now *your* turn to speak. To be able to respond appropriately and continue the conversation you need to be able to estimate the duration of the pause, a judgment in the millisecond to second time range, and compare it to an internal standard in order to decide whether you should now say something. An accurate estimate could help you to avoid social embarrassment, the so-called "awkward silence" (Burgoon & Saine, 1978). This is but just one simple example demonstrating the importance of duration judgments in human perception and behaviour. From simple characteristics, such as the duration of such a pause in speech, to anticipatory behaviours (i.e., estimating and predicting differences in duration between consecutive intervals), temporal estimates in the millisecond-to-second range guide our perception of, and interactions with the environment (Buhusi & Meck, 2005; Buonomano, & Karmarker, 2002; Karmarker & Buonomano, 2007; Ivry & Schlerf, 2008). Despite the importance of these estimates, there are no dedicated receptors for time and duration can only be estimated from variations occurring in signals themselves once two of these variations (e.g., on- and off-set) have already occurred. Thus, any pair of signal changes can give rise to a duration estimate, may they be visual, haptic, or auditory. Each of the two changes can also be derived from concurrent signals activating different sensory modalities (Burr et al., 2009a; Grondin, 2001; Grondin, Oullet, & Roussel, 2001; Ulrich et al., 2006; van Wassenhove et al., 2008). In this chapter, the integration of duration estimates was investigated when these changes across the signals occur in vision and audition.

Many studies have investigated the integration process of redundant multisensory signals. For example, the estimate of spatial properties such as location, shape, or size has been shown to result from a weighted averaging of the individual estimates (Alais & Burr, 2004; Clarke & Yuille, 1990; Ernst & Banks, 2002; Ernst & Büthoff, 2004; Helbig & Ernst, 2007; Hillis, Watt, Landy, & Banks, 2004). This can be expressed mathematically in terms of the following formulas:

$$S = \sum_i w_i S_i \tag{7.1}$$

where the weights are proportional to reliability of the estimates according to:

$$w_i = \frac{r_j}{\sum_{i=1\ldots j\ldots N} r_i} \tag{7.2}$$

The reliability *r* in Equation 7.2 is defined as the inverse variance of the estimates (see Ernst & Bülthoff, 2004, for a review):

$$r_i = \frac{1}{\sigma_i^2} \tag{7.3}$$

With such weights, the integrated estimate has the lowest possible variance (highest possible reliability *r*) according to:

$$r = \sum_i r_i \tag{7.4}$$

The increased reliability of the combined estimate with respect to the most reliable unisensory estimate is the benefit of integration. When the increase in reliability is maximal as specified by Equation 7.4, the integration is said to be 'statistically optimal' (e.g., Clarke & Yuille, 1990). Integrating signals in this way is also referred to as Maximum-Likelihood-Estimation (MLE). Optimal integration has been demonstrated for several types of spatial estimates (i.e., Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Knill & Saunders, 2003). Moreover, optimal integration has also been observed in the temporal domain (Ley et al., 2009). Ley et al. demonstrated that when observers were asked to indicate on which side the signal occurred first, temporal order estimates of two multisensory events were indeed optimal relative to the unisensory conditions. For multisensory duration estimates, however, it is still unknown whether humans integrate information optimally.

At this point, one may ask why the integration of duration information might be different from all the other cases where optimal integration has been observed? Spatial estimates, and also the estimate of a multisensorially-defined time point used for TOJs (Ley et al., 2009) can be integrated 'online' while the information is still available sensorially. Estimating the duration of an event is distinctly different because duration can be defined only once the event has already ended. Thus, when estimates have to be made and integrated there is no longer any sensory information available anymore. In that sense, then, duration estimates are post-hoc (Grondin, 2001).

It has been hypothesised that this might be the reason why, in the literature, multisensory duration estimation has been shown to be suboptimal (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011). For example, Burr et al. observed auditory driving, that is, the integrated percept of duration to be dominated by the auditory modality and it did not follow the prediction of Equation 7.1. Now the question is whether this is generally true for duration estimates or if this result depends on the experimental details of the studies concerned? In Burr et al.'s study, temporal intervals were defined by the time elapsed between two short stimuli (markers) which could be auditory, visual, or audiovisual. No signal was present during the judged interval itself, but only at the onset and offset of the interval. Such a stimulus is defined as an "empty interval" (Allan & Kristofferson, 1974; Grondin, 1993, 2001). This stimulus type is ambiguous regarding which temporal property is redundantly specified and thus undergoes multisensory integration. A participant could estimate the duration of unimodal intervals separately and subsequently integrate them into a unified percept (this case is defined as "redundant duration"). Alternatively, however, they could first combine the audio and visual markers and then estimate the duration between these two integrated time points (this is the case of "redundant time points"). The mismatch between the observed data and the predictions from the MLE model might have been based on this difference: The predictions are based on the assumption that visual and auditory duration estimates are integrated. However, the empirical integration results could have been derived from estimates of the integrated markers.

In order to remove the possibility that onset and offset markers could be integrated before they are used for estimating duration, the stimuli used in the experiment reported in this chapter are audio, visual or audiovisual "filled interval" (Allan & Kristofferson, 1974) as represented in Figure 7.1b. In this manner the signal is present throughout the entire interval rather than only at the boundaries. If such a confound is removed, and the pure integration of duration estimates is examined, will integration be close to optimal?

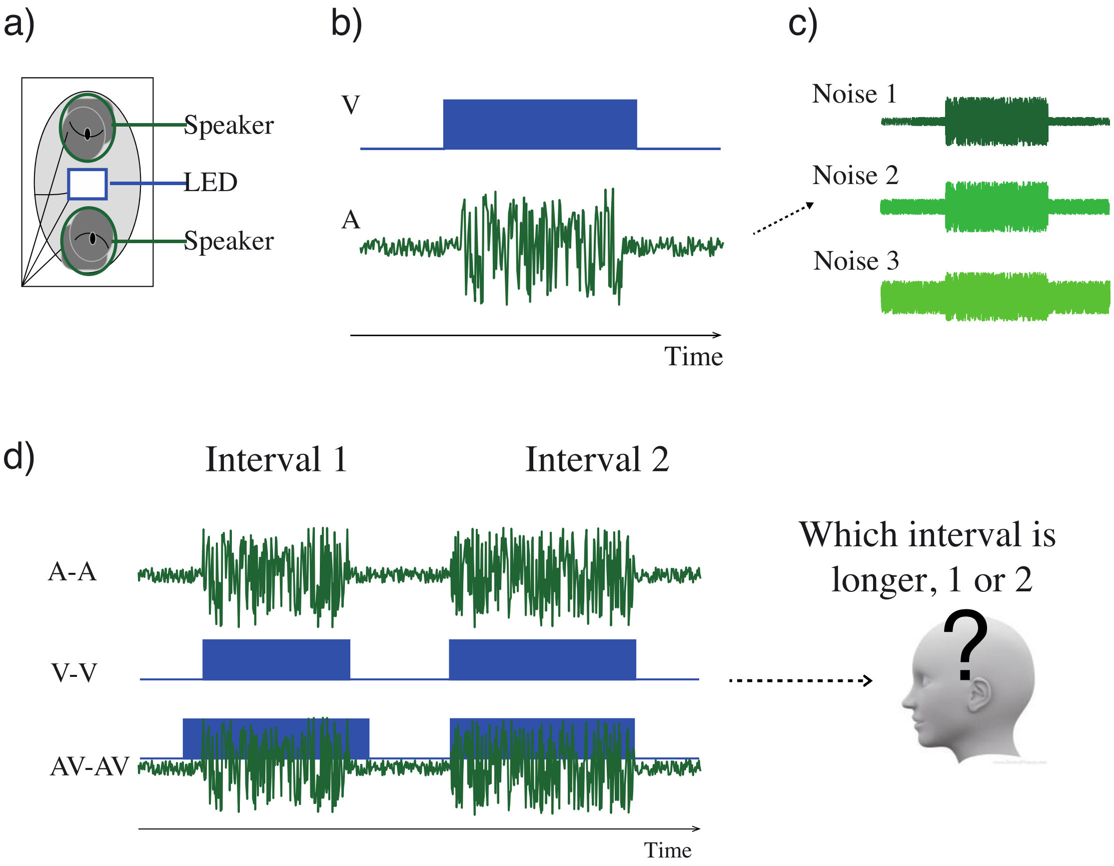


Figure 7.1. a) The stimuli were presented from co-located loudspeakers and LED array. b) Schematic representation of audio (green) and visual (blue) signals used in Experiment 7.1 c) The signal-to-noise ratio of the auditory signal is manipulated in three levels. d) 2IFC procedure – which interval is longer. Each trial is comprised by a standard and a comparison duration presented in a random order.

7.2 Methods

7.2.1 Observers

Eleven volunteer observers, all reporting normal hearing and normal or corrected-to-normal vision participated in the experiment. Observers gave their informed consent prior to their inclusion in the experiment, they were naive to the purpose of the experiment, and they received a compensation of 8 Euros per hour of participation. The Ethics committee of the Eberhard Karls University of Tübingen gave approval for the study. Three observers' data sets have been excluded from the analysis due to low discrimination performance in the uni-sensory conditions (average Weber Fractions across all conditions for these observers exceeded three standard deviations from the mean of all observers' Weber Fraction and was thus close to chance performance), leaving 8 observers datasets.

7.2.2 Apparatus

Stimuli were produced using a custom-built device that can generate co-located sound and light signals. Observers sat with their head approximately 60 cm from the device, which looked like a box made of wood and foam standing on a table. Two 2.5 cm radius vertically aligned speakers with a centre-to-centre distance of 7.5 cm were concealed behind foam and produced the auditory stimuli. An LED matrix display (7 x 5 red LEDs, 1.6 x 1.3 cm) was mounted between the loudspeakers and visible to the participant through a hole in the foam. A multi-channel sound card (M-audio 1010LT) and two identical audio amplifiers were used to generate the stimuli, ensuring synchronization between the signals. The experiment was conducted in a dimly lit, sound attenuated room.

7.2.3 Stimuli

Auditory signals were broadband noise where the peak intensity was 61 dB SPL. Visual signals from the LED array were 150 Hz sinusoids where negative values were rectified to also cause the LEDs to light up. The luminance of the stimulus was 41 cd/m^2. The average duration of the signals across all trials was 500 ms. Signals had a 5 ms linearly ramped onset and offset.

In order to alter the reliability of the auditory signals, the auditory noise burst (signal) was embedded in continuous background noise and its intensity was manipulated (0.1, 0.6 and 1.2 times the signal level). The auditory background noise was continuously presented throughout a trial with a random time span between 200 and 450 ms before and after stimulus presentation. No noise was added to the visual signal. In Figure 7.1, the auditory (green) and visual (blue) signal pairs used in the experiment are represented schematically.

7.2.4 Procedure

A two-interval, forced-choice procedure was used. Each trial consisted of the sequential presentation of two intervals either auditory or visual or audiovisual and observers required to indicate which interval perceptually lasted longer (see Figure 7.1c). The inter-stimulus interval (ISI) varied randomly between 500 and 900 ms. The duration of the standard interval in the uni-sensory trials (i.e. audition alone or vision alone) could be one of three durations: 450 ms, 500 ms, or 550 ms; and for the comparison interval, the duration could be ± 5, 10, 20, or 40% of the standard interval in that trial, varied according to the method of constant stimuli. The order in which the standard and comparison were presented was randomized. The observer's task was to indicate which interval appeared to last longer.

In the multisensory (audiovisual) trials, observers compared two intervals for which the visual and auditory signals (with one of three noise levels) were present in both intervals. In these trials the procedure was identical to that implemented in the uni-sensory trials, only the duration of the audio and visual stimuli of the standard interval contained a small discrepancy and thus could be different by Δ=0, -50, +50 ms. The signals were aligned at the temporal midpoint such that the discrepancy was distributed equally at either end of the stimulus. Trials with standard intervals having different discrepancies were presented interleaved in random order.

Each comparison was repeated 16 times giving a total of 5616 responses per subject (702 in each of the 8 experimental runs each lasting about 30 minutes). Responses were fit with a cumulative Gaussian psychometric function using the Psignifit toolbox version 2.5.6 for Matlab. The toolbox implements the Maximum likelihood model as described by Wichmann and Hill (2001). From the cumulative Gaussian fits, the points of subjective equality (PSE) and the just noticeable differences (JNDs) were obtained. The PSE corresponds to the duration at which the proportion of responses indicating the comparison appears longer reaches the 0.50 level – thus comparison and standard interval are perceived as equally long. The JND corresponds to the difference in duration between standard and comparison interval that is necessary to increase discrimination performance from 0.50 to 0.84. The Weber Fraction (WF) was calculated for each condition such that $WF=JND/D_S$ where D_S is the duration of the standard interval. In this way it was possible to collapse across the three standard durations using the median. For audiovisual stimuli, we took the mean between D_{SV} and D_{SA} as a measure of D_S. Thus, the WF indicates the precision of the duration judgment in the discrimination process. Trials with uni-sensory and audiovisual stimuli were interleaved throughout the experiment. From the unisensory trials predictions were derived for optimal performance in the multisensory condition according to the MLE equations (in Equations 7.1-7.4). This made it possible to compare the empirical performance in the estimate of duration based on audiovisual stimuli with the one predicted from the performance obtained with unisensory stimuli. In the following section the unisensory results will first be described, followed by the comparison between prediction and empirical findings for the audiovisual trials.

7.3 Results

7.3.1 Unisensory Weber Fractions

Figure 7.2a depicts unisensory visual and auditory WF data for observer MDJ, a representative observer whose data will be given as example throughout this Chapter. In three conditions, auditory noise was added to the background (low, medium, or high noise levels, see Figure 7.1b), thus modulating the signal-to-noise ratio, but the visual condition was always noise free. For auditory stimuli, duration discrimination thresholds increase systematically with increasing noise levels. Visual estimates are approximately as precise as the auditory estimate in the middle noise condition (see Figure 7.2a). Uni-sensory WF data for all observers are presented in the histograms of Figure 7.2b-e. The mean auditory WF goes from 19% to 54% as the noise level of the auditory stimulus increases (one-way Repeated Measures (RM) ANOVA $F(2,7)=35$ $p<0.001$). The average visual WF is 29%.

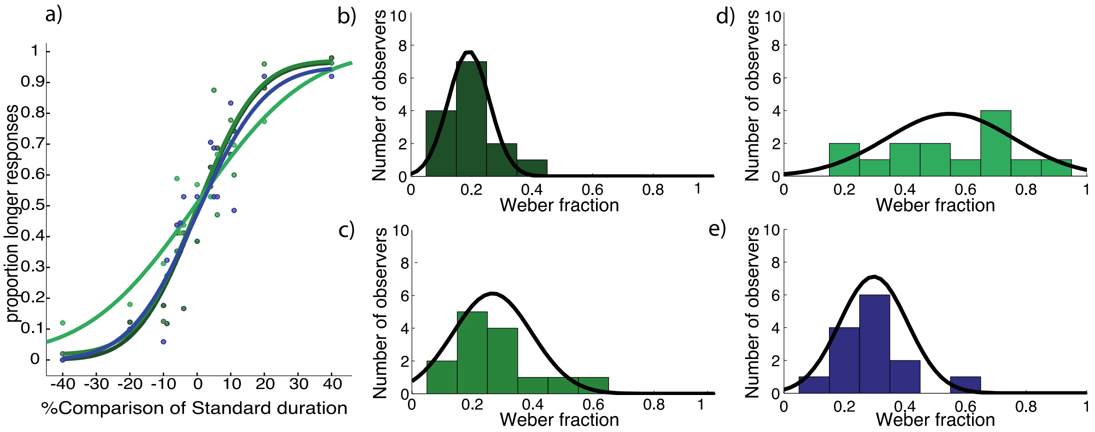


Figure 7.2. a) Example observer MDJ's unisensory psychometric functions for the three auditory noise levels (the shade of green indicates the level) and vision (blue). b, c, d) Distribution of Weber fraction values across the 8 observers for the auditory condition ordered for noise level. e) Distribution of Weber fraction values for the vision condition. All data are from Experiment 7.1.

7.3.2 Multisensory WF: Predictions and empirical findings

Individual WFs in the unisensory conditions were used to predict performance under multisensory conditions according to the MLE model. For this, Equations 7.1-7.4 are applied to the integration of redundant duration information, indicating that perceived duration of audiovisual stimuli D_{AV} should be a weighted average of the auditory and visual components according to the following formula:

$$D_{AV} = w_V D_V + w_A D_A \qquad (7.5)$$

where the weights of the unisensory estimates of duration are calculated using the individual WF according to:

$$w_V = \frac{WF_A^2}{WF_A^2 + WF_V^2} \quad \text{and} \quad w_A = \frac{WF_V^2}{WF_A^2 + WF_V^2} \qquad (7.6)$$

These weights lead to a maximal decrease in uncertainty for multisensory estimates of duration such that the Weber fraction in the bimodal conditions can be calculated from the unimodal ones:

$$WF_{AV}^2 = \frac{WF_V^2 \, WF_A^2}{WF_A^2 + WF_V^2} \leq \min(\mathrm{WF}_V^2, \mathrm{WF}_A^2) \tag{7.7}$$

Equation 7.7 predicts a reduction of WF (and correspondingly a reduction in variance) for the multisensory duration estimate compared to the best unisensory estimate. This prediction, one of the signatures of optimal integration, can be verified for observer MDJ in Figure 7.3a, where predicted values are shown alongside empirical WFs, obtained from the unisensory and multisensory conditions. In Figures 7.3b and 7.3c, the same comparison can be made for data averaged across observers. The mean empirical multisensory WF across the three noise levels has a similar pattern to the predictions obtained with Equation 7.7 (see Figure 7.3a). A 3 (noise level) times 2 (empirical vs. predicted) RM 2-way ANOVA on the WF values does not indicate significant deviations from predictions for empirical values alone (factor type of WF: $F<1$, n.s.) nor in conjunction with the noise level (interaction term: $F<1$, n.s.). Performance in the bimodal condition is also well predicted by MLE at the individual level as evidenced by the similarity between predicted and audiovisual observed WF (see Figure 7.3c). A regression line fitted to the data has a slope of 0.76 (95% C.I. 0.58-0.94) and an intercept of 0.007 (95% C.I. 0.02-0.13) with $R^2=0.78$ ($p<0.001$) indicating that the MLE model successfully predicts the individual performance improvement due to multisensory integration. Moreover, the MLE model predicts that the combined cue estimate is more precise than the best unisensory WF estimate. A 2*3 RM ANOVA comparing condition (best unisensory and multisensory WF) and noise (three auditory noise levels) revealed no significant main effect of condition, F(1, 7)=4.86, *p*=0.63. Thus the difference between best unisensory and observed WF was not significant overall. However, the main effect of noise was significant, F(2, 7)=14.49, *p*<0.01. Furthermore there was also a significant interaction between condition and noise level, F(2, 14)=5.31, *p*=0.02. Given the significant interaction, post-hoc comparison paired-sample t-tests were performed to understand the nature of the interaction. These revealed that the best unisensory WF estimate is significantly higher than the multisensory WF for the intermediate noise level (one-tailed paired-sample t-test: t(7)=3.7 *p*=0.007). In this condition the precision of the visual and auditory duration estimates are comparable. As such, bimodal estimates have the largest predicted and observed increase in reliability when compared to the unisensory estimates. For the data reported here it was precisely under these conditions that unequivocal, optimal integration was observed. In order to quantify the support for the hypothesis that there is no difference between MLE predictions and the observed multisensory WFs, the JZS Bayes Factor was calculated (see Gallistel, 2009). The probability that the null hypothesis as opposed to the alternative difference hypothesis is supported is translated from the likelihood ratio to a Bayes Factor of 3.58.

Not surprisingly, in the two extreme noise conditions the difference in reliability between the auditory and the visual duration estimates is substantial, the predicted improvement in the bimodal condition is small in magnitude, and the reliability of the multisensory estimates is thus not statistically different from the best unisensory estimate t(7)=1.4 *p*=0.19; and t(7)=1.9 *p*=0.10 for the lowest and highest noise level conditions respectively. Importantly, however, under these extreme noise conditions, the observed multisensory WFs are also not different from the optimal predictions. In summary, the decrease in variance associated with audiovisual duration estimates observed especially in the intermediate noise condition is one of the signatures of optimal integration and thus provides the first evidence that optimal integration occurs for multisensory duration estimation.

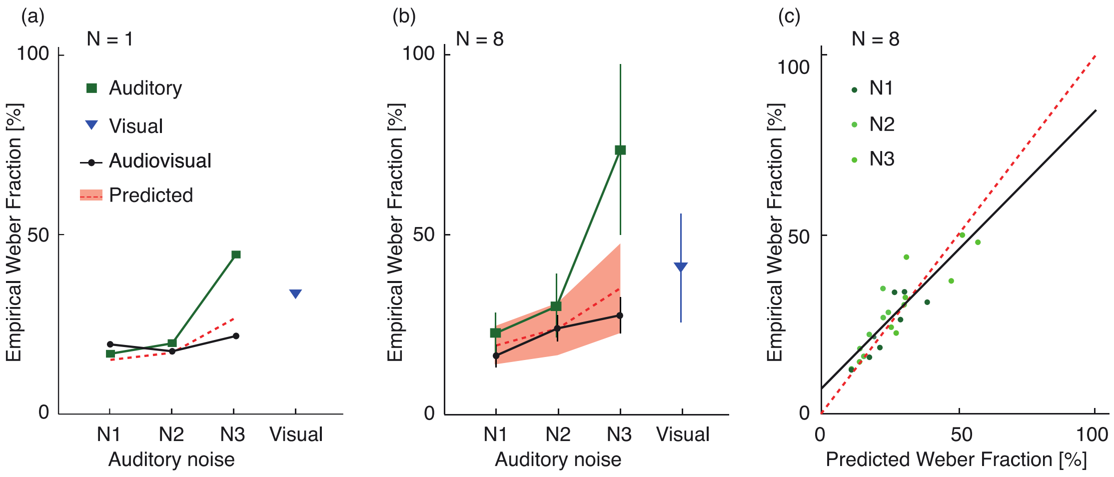


Figure 7.3. a) Example observer MDJ Weber Fraction values for unisensory and, multisensory conditions and MLE predictions. b) Mean unisensory, multisensory, and MLE predicted WF values across observers. Unisensory WF data is obtained from the distributions represented in Figure 7.2b-e. Predicted values are instead obtained from Equation 7.7. c) Relation between empirical and predicted Weber fraction values across observers. For optimal integration the mapping between observed and predicted should be a 1-to-1 relationship. The line of best fit is consistent with such a mapping. All results are for Experiment 7.1.

7.3.3 Multisensory weights: Predictions and empirical findings for conflict conditions

MLE predicts that in order to maximally decrease the influence of noise on a perceptual estimate, integration should be a function of the weights assigned to the redundant estimates (as stated in Equation 7.6). According to MLE predictions, as the noise in the auditory signal increases, the weight w_v assigned to the visual component of the estimate $\widehat{D}_V$ should increase. The values of PSE$_{AV}$, that represent the perceived duration of audiovisual stimuli containing a temporal conflict, can be used to calculate the empirical weights assigned by each observer. A comparison between predicted and empirical weights at the different noise levels is shown in Figure 7.4a for subject MDJ and in Figure 7.4b for the average across the group of all observers. It is evident that, with increasing auditory noise (see Figure 7.4a & b) the estimate of duration relies more and more on visual information (the weight given to the visual component increases). The empirical pattern of individual weights was obtained by applying Equation 7.5 to the values of PSE_{AV} at each noise level. These values nicely follow the predicted weights derived from Equation 7.6, as indicated by the correlation between predicted and empirical weights visible in Figure 7.4c ($R^2=0.65$, $p<0.001$). There is no significant difference between empirical and predicted weights across observers, given that the regression line does not deviate statistically from the diagonal (slope of 1.1, 95% C.I. 0.77-1.5 and intercept of -0.07, 95% C.I. -0.02-0.10).

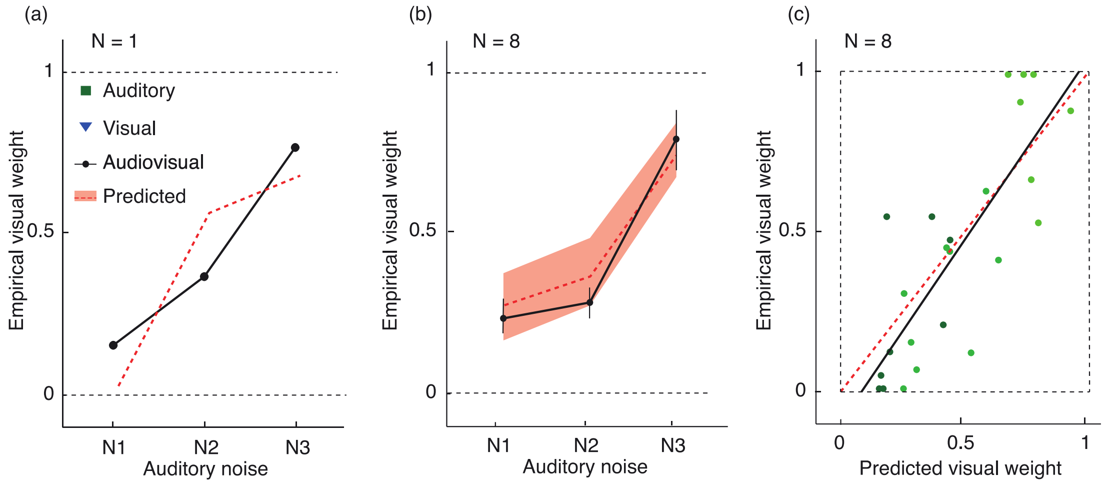


Figure 7.4. a) Example observer MDJ's values of PSE_{AV} in multisensory conflict conditions expressed in terms of visual weight. MLE predictions indicate that as the noise in the audio signal increases the visual weight should increases correspondingly. b) Average values of visual weight in multisensory conflict conditions. c) Individual visual weights showing the correlation between empirical values and predictions for the three noise conditions. The regression line shows the mapping between the predicted and observed weights. As in Figures 7.2 and 7.3, the data presented here are from Experiment 7.1.

Taken together, the results for the Weber Fractions and the weighting behaviour indicate that observers seem to optimally weight the audio and visual components to obtain a single estimate of multisensory duration.

7.4 Discussion

The experimental work described in this chapter examines how the nervous system obtains an integrated duration estimate of intervals redundantly specified by multisensory signals. Whether such an integrated estimate is statistically optimal, despite the fact that the perceived duration of an event can only be obtained post-hoc, i.e., once the sensory information is no longer available was investigated. The crucial finding is that redundant audiovisual duration information is integrated in a statistically optimal fashion. That is, the results demonstrate that the multisensory duration estimate is a weighted average of the unisensory estimates with weights proportional to their reliabilities. Moreover, when the duration signals are integrated, there was an increase in reliability for the multisensory estimate corresponding to the predictions of the MLE model.

Here, for the first time, the integration of redundant estimates of the temporal extent of an audiovisual event is shown to be close to optimal. Optimality has previously been demonstrated in the integration of spatial estimates (Alais & Burr, 2004; Ernst & Banks, 2002; Helbig & Ernst, 2007; Jacobs, 1999), but also in the perception of temporal order (Ley et al., 2009). Together with the findings presented in the present chapter, Ley et al.'s results demonstrate that redundant temporal information is integrated optimally. This contrasts with earlier suggestions that temporal duration information is driven primarily by the auditory input (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011). Given this conflict between findings, the question therefore arises as to whether there may be several ways in which the nervous system can deal with redundant temporal information for the estimate of duration. To answer this question, let us consider what information is available to estimate duration in relation to the stimuli that have been used in the different experimental investigations of multisensory duration judgments. All of the studies that have looked at multisensory integration in the time domain previously have used short markers – auditory or visual stimuli of brief duration (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011; Ley et al. 2009). With such short stimuli, in the bimodal conditions, the multisensory duration estimates could be obtained in two ways. The nervous system may either obtain an estimate of duration for the audio and visual channels independently. The two redundant estimates of duration would then be integrated into a unified estimate of multisensory duration. Similar to many other studies of multisensory integration (e.g., Alais & Burr, 2004; Deneve & Pouget, 2004; Ernst & Banks, 2002; van Beers et al., 1999) predictions for human performance can be made following Equations 7.5-7.7. This case is the "redundant duration information", named so as to distinguish it from the alternative case "redundant time point information". Alternatively, with short multisensory stimuli demarking the intervals, it could be argued that the markers are used to obtain an integrated estimate of the start and end of the interval. Duration judgments would then be made based on these integrated markers and not by integrating two redundant duration estimates. In such a case, predictions cannot be made using Equations 7.5-7.7, rather they follow a different scheme as described below.

Let us now compare the performance that can be achieved in duration judgments with the two proposed computational models that involve either integration of duration information or of the interval markers' time points. In either case, the precision of a duration estimate is limited by two noise sources, one due to the markers (both on- and off-set) and one source is instead due to the storage of temporal information concerning the onset until the interval is ended by the offset. The latter component represents a noise source in the duration estimate that comes from memory and which makes longer durations more difficult to discriminate than short ones. The variance σ_D^2 of a duration judgment can then be expressed as:

$$\sigma_D^2 = \sigma_{ON}^2 + \sigma_{OFF}^2 + \sigma_S^2 = 2\sigma_M^2 + \sigma_S^2 \tag{7.8}$$

where it is assumed that the variances associated with the on- and off- set are equal ($\sigma_{ON}^2 = \sigma_{OFF}^2 = \sigma_M^2$). If the interval is defined by multisensory signals and the duration information is redundant and thus integrated, the variability of the integrated estimate $\overline{\sigma_{D,AV}^2}$ can be expressed as a function of the variability of the unisensory estimates $\sigma_{D,A}^2$ and $\sigma_{D,V}^2$ following Equation 7.7:

$$\overline{\sigma_{D,AV}^2} = \frac{\sigma_{D,A}^2\,\sigma_{D,V}^2}{\sigma_{D,A}^2 + \sigma_{D,V}^2} \tag{7.9}$$

The single line superscript indicates that the overall duration judgment is obtained through the integration of redundant duration information. Substituting the unisensory variances $\sigma_{D,A}^2$ and $\sigma_{D,V}^2$ with Equation 7.8 (and assuming that σ_S^2 is equal in the two modalities) expands to:

$$\overline{\sigma_{D,AV}^2} = \frac{\left(2\sigma_{M,A}^2 + \sigma_S^2\right)\left(2\sigma_{M,V}^2 + \sigma_S^2\right)}{2\sigma_{M,A}^2 + 2\sigma_{M,V}^2 + 2\sigma_S^2} \tag{7.10}$$

The same substitution can be performed for Equation 7.6 in order to obtain the weight assigned to the unisensory estimates as a function of markers and storage noise:

$$\overline{w_A} = \frac{1/\left(2\sigma_{M,A}^2 + V_S\right)}{1/\left(2\sigma_{M,A}^2 + V_S\right) + 1/\left(2\sigma_{M,V}^2 + V_S\right)} \tag{7.11}$$

If instead the audio and visual markers are integrated and duration is estimated on such time points, the variability of the on- and off-set markers becomes:

$$\sigma_{M,AV}^2 = \frac{\sigma_{M,A}^2\,\sigma_{M,V}^2}{\sigma_{M,A}^2 + \sigma_{M,V}^2} \tag{7.12}$$

and the weight given to the two markers is instead a function of the marker's reliabilities according to:

$$\overline{\overline{w_A}} = \frac{1/\sigma_{M,A}^2}{1/\sigma_{M,A}^2 + 1/\sigma_{M,V}^2} \tag{7.13}$$

The two-line superscript indicates that the duration judgment is made between two time points where the audio and visual markers are integrated. This leads to a variability of the duration estimate of audiovisual conditions that is expressed by:

$$\overline{\overline{\sigma_{D,AV}^2}} = 2\frac{\sigma_{M,A}^2\,\sigma_{M,V}^2}{\sigma_{M,A}^2 + \sigma_{M,V}^2} + \sigma_S^2 \tag{7.14}$$

It is easy to demonstrate that if the storage of temporal information does not cause a decrease in performance for the overall duration judgment (if σ_S^2=0), then there is no difference in the two methods of integration, either in terms of variability (Equation 7.10 is equal to Equation 7.14) or weighting of the audio and visual components (Equation 7.11 is equivalent to Equation 7.13). However, if $\sigma_S^2 > 0$, then the variability of duration judgments based on integrated markers described by Equation 7.14 is always greater than that obtained through the integration of redundant duration information expressed by Equation 7.10. The reason for this imbalance is that when the time points of the multisensory markers are treated as redundant and the duration estimate is derived from an integrated time point, the noise term associated with the memory storage component σ_S^2 appears as an additional factor in the equation of the variance (Equation 7.14), but it is not considered in the weighting (Equation 7.13) leading to suboptimal integration. In fact, since the variances with which the onset and the offset maker can be determined may well differ from the variance of the duration estimate because of the additional memory storage term. That is, the variance of the duration estimate is determined by more than just the variance of the onset and offset markers. There must be an additional noise source in the duration estimate from memory (or a comparable component), which makes longer durations more difficult to discriminate. However, the variance in determining the point in time for the on- and offset, respectively, is independent of the duration that passes between those markers. As a consequence, if the variance associated with the two modalities is not equal, the weight assigned to the most reliable component (in the time domain, this is normally the auditory modality) is higher for the model where the time points are integrated than for the model where duration information is integrated. In other words, integrating information from markers rather than for duration would be associated with the overweighting of the auditory modality and an associated higher variance for the overall duration judgment. This might explain the apparent contradiction between the current results showing optimal integration and those of Burr et al. (2009a) revealing an overweighting of auditory information. It should also be noted that the current study uses filled intervals – intervals are defined by the presence of a continuous visual, auditory, or bimodal stimulus. Integration at the level of markers' time points with this kind of stimuli is difficult to conceive, and it is the possibly the reason why optimal performance was predicted and observed here whereas others who worked with empty intervals were unable to demonstrate this.

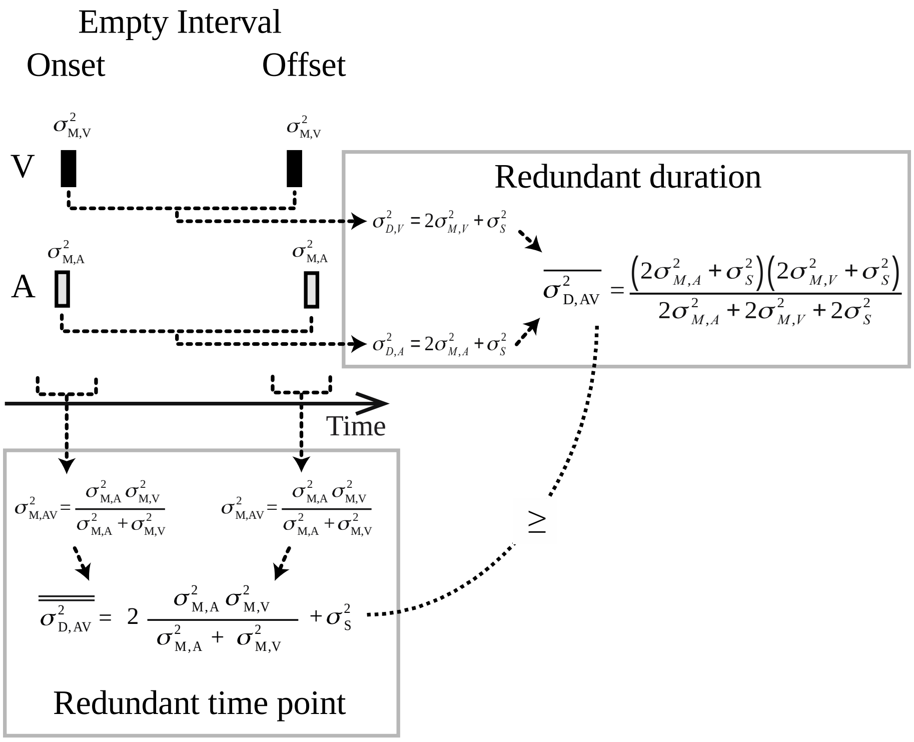


Figure 7.5. Depiction of a single audiovisual empty interval (i.e., as used by Burr et al., 2009a). Short audio and visual on-set and off-set markers delineate the interval whose duration is to be estimated. If observers integrate redundant unisensory estimates of the interval duration, this would lead to the prediction given in the box titled "redundant duration". However, it is also possible that observers integrate the unisensory estimate of time for the on-set and off-set markers giving rise to the box titled "redundant time point". In this second case, the audio and visual markers are first integrated and only at a later stage the duration estimate is made on the integrated markers.

The most compelling evidence for optimal integration comes from those conditions in which there an is equal weighting of the unisensory estimates, in which case the bimodal estimate has the largest increase in reliability when compared to the unimodal ones. For the data reported in this chapter, it was precisely under these conditions that unequivocal optimal integration was observed, the condition with intermediate auditory noise. In the two extreme noise conditions the reliability of the multisensory estimates is not statistically different from optimal predictions, but also not different from the best unisensory estimate either.

The current findings provide insights into the strategy underlying redundant cue integration for *perceived* event duration (cf. Burr et al., 2009a). Although recent findings highlight the importance of models of duration reproduction (see Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012; Jazayeri & Shadlen, 2010), perceived duration has remained ambiguous until now. For duration reproduction, it has been suggested that both humans (Cicchini et al., 2012; Jazayeri & Shadlen, 2010) and rats (Balci, Freestone, Simen, deSouza, Cohen, & Holmes, 2011) are able to learn the temporal statistics and optimally reproduce durations in the target sub-second distribution. Jazayeri and Shadlen (2012) instructed their observers to reproduce time intervals that were sampled from different underlying distributions (including sub-second intervals). When reproductions fell within a temporal window of the target interval, observers received positive feedback. The resulting reproductions of target intervals were observed to regress to the mean. Most importantly, a cost-minimization model that takes into account the statistics of the target duration distributions can describe the results. These findings have been more recently confirmed by a study with human musicians (Cicchini et al., 2012)

The results from Cicchini et al. (2012) and Jazayeri and Shadlen (2010), coupled with the current findings, provide evidence for a nervous system that deals optimally with temporal information. When observers are asked to estimate a single moment of audiovisual event duration they utilize the redundant cue information to minimize the uncertainty in their estimate and, furthermore when asked to reproduce a duration they draw from a prior distribution of durations to approximate the target signal. Yet, it has been suggested (Buhusi & Meck, 2005) that duration perception in the ms-sec range and reproduction of duration are the result of distinct underlying mechanisms. However it appears that for both, observers use the statistics of the signals themselves and of the history of events to minimize uncertainty. Furthermore, for MLE to successfully describe the integration process an assumption of independent channels needs to be satisfied (Oruc, Maloney, & Landy, 2003) and this has important implications for duration estimation.

What makes optimal integration for perceived duration so intriguing is that it can provide some insight into current debates concerning how mechanisms of timekeeping are implemented in the nervous system. Duration estimation has been used to understand how temporal information is coded and processed in humans and animals alike (Buhusi & Meck, 2005; Matell & Meck, 2000). The prevailing model for event timing involves a centralized internal clock, or pacemaker, feeding into an accumulator (for example, see Gibbon, 1977; Gibbon & Church, 1984; Treisman, 1963). More recent models, however, consider distributed timing networks, with different mechanisms timing different interval lengths (Buhusi & Meck, 2005; Burr, Tozzi, & Morrone, 2007; Heron et al., 2012b; Karmarker & Buonomano, 2007). While there is good evidence for different clocks for different interval lengths, most evidence to date suggests that clocks transcend sensory modalities. For example, after adaptation to a particular duration of an auditory signal, perceived duration of subsequent auditory intervals is compressed. The same effect is found for subsequent visual duration estimates (Walker, Irion, & Gordon, 1981), implying a common mechanism across sensory modalities. However, more recent findings suggest, in fact, that there seems to be timekeepers dedicated to duration discrimination in each modality (Burr et al., 2007; Heron et al., 2012b). Heron et al. suggested that there are multiple channels of duration discrimination based on different signal duration channels. They proposed a channel model to account for adaptation effects that were specific for different duration bands. Together with the present findings and those of others (Alais & Cass, 2010; Burr et al. 2007; Wiener, Matell, & Coslett, 2011), the evidence is, then, now mounting for multiple timekeepers. Moreover, duration estimation itself differs from all the other dimensions investigated for redundant cue combination so far (for which optimal integration has been found) in that it is a post-hoc estimate. That is, the perceptual judgment cannot be made while the sensory information is still available. Duration can only be judged when the event has already been experienced. Being optimal in integrating multisensory event duration therefore means that the integration mechanism operates on the stored representation of event duration and its associated variance. With MLE, there is the assumption of independent channels, therefore optimal integration of event duration suggests that the mechanism responsible for duration perception is a multiple channel mechanisms given estimates must be independent.

In this chapter, it has been argued that integrating multiple cues into a coherent duration estimate is a particularly interesting process because of its post-hoc nature which sets it apart from several other forms of multisensory integration where optimality has been established – i.e., location, size, orientation etc. The findings from the current experiment provide the first evidence that audiovisual duration is optimal as described by the MLE model. Being optimal in integrating multisensory event duration therefore means that the integration mechanism operates on the stored representation of event duration and its associated variance. For MLE to successfully describe the integration process an assumption of independent channels needs to be satisfied (Oruc, Maloney, & Landy, 2003), therefore optimal integration of event duration provides a hint that duration estimation may be obtained through an independent channels mechanism (i.e., as suggested by Heron et al., 2012b; Wiener, Matell, & Coslett, 2011).

7.5 Conclusions

The nervous system can integrate redundant information about the temporal extent of a multisensory event so as to reduce the uncertainty in its perceptual estimate of duration. The performance is close to optimal when information about unisensory duration is integrated, in contrast to the case when the time points defining the interval are integrated. For duration estimates the nervous system integrates redundant information in a manner similar to spatial estimates, despite the fact that the information is no longer available when the estimate is made. So, next time you estimate an interruption in speech and you want to avoid that awkward silence, watch AND listen!

Chapter 8: General discussion

8.1 Summary of results

The experiments reported in this thesis explored how complementary and redundant cues are used by the nervous system in the processing and integration of multisensory information that specifies spatial and temporal properties of the environment.

In the study described in Chapter 2 observers were presented with a visual flash, a tactile vibration to the fingertip or both cues simultaneously. Their task was to respond as quickly as possible to visual flash or vibration or, when presented simultaneously, to both.

Responses to the tactile pulse and visual flash were highly accurate. Responses to the two cues presented simultaneous on the other hand, were not: Observers failed to respond to the tactile pulse cue on a significant number of these multisensory trials. This visual dominance for detecting the flash and inhibiting the tactile pulse was stronger for stimuli at threshold (Experiment 4). When the probability of multisensory cues was increased, the visual dominance decreased (Experiment 2).

In the experiment described in Chapter 3, observers performed the same speeded detection task to flashes and tactile pulses. However, 33% of visual flashes were either presented to one of three locations: the observers fingertip, a distant location in extrapersonal space, or on a rubber hand at the distant location in extrapersonal space.

Unisensory visual responses to the simultaneously presented flashes and tactile pulses occurred when the visual flash was presented on the observer's body or on the rubber hand although it was located in extrapersonal space. The visual dominance effect did not transfer to extrapersonal space when the rubber hand was not present. These results therefore suggest that the competition between the complementary flash and tactile pulse is strongest for events on a body surface.

For the study discussed in Chapter 4, the influence of short-term temporal sequences on responses was explicitly tested. Observers were exposed to a continuous train of redundant but slightly discrepant beeps and flashes. After this exposure period, they were required to respond as rapidly as possible whenever they detected unisensory beeps or flashes of light.

RTs to these unisensory signals changed as a result of the type of discrepancy presented in the exposure phase. If the beeps preceded the flashes of light then the consequent unisensory auditory RTs were slower than when the beep lagged behind the flash in the exposure period. This change in auditory RTs is consistent with the nervous system recalibrating to the temporal discrepancy between the beep and flash.

In the study reported in Chapter 5, observers were again exposed to discrepant beep and flash sequences. But this time exposure was intermittent. After 40 seconds of being presented with the discrepant beeps and flashes, observers were provided with unisensory beeps and flashes and required to detect and respond as quickly as possible to these unisensory signals. The change in the RTs to the unisensory beeps and flashes was measured over time.

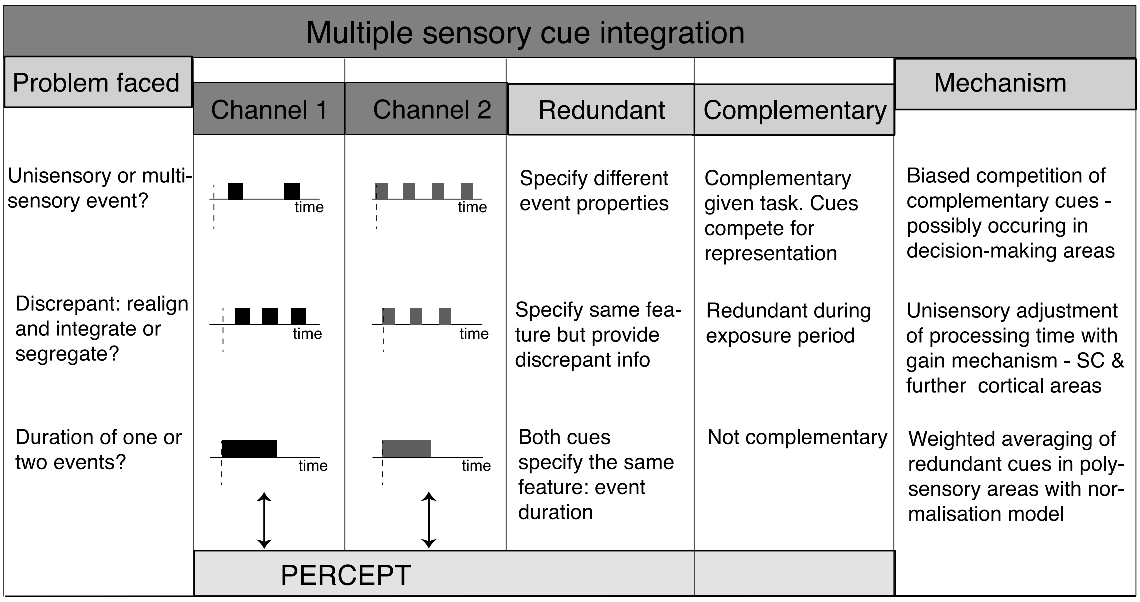
However, neither the auditory nor the visual RTs shifted after the intermittent exposure to the discrepant beeps and flashes. The lack of observed recalibration highlights an important pre-requisite for temporal recalibration to occur: exposure to the discrepancy needs to be continuous. Otherwise the cues, though redundant, are segregated and treated as arising from two distinct events.

The work reported in Chapter 6 explored a slightly different aspect of recalibration: perceived simultaneity of the beep and flash after exposure. Following continuous and prolonged exposure a beep of one auditory frequency and a flash occurring at a slightly different time, observers made SJ to single beep-flash pairs. The beep-flash pair could either contain the exposed beep frequency or a beep frequency not presented before. Perceived simultaneity only changed for the adapted auditory frequency. The un-adapted frequency signal was inferred to arise from an event separate to that to have undergone recalibration.

From these temporal recalibration studies it can be concluded that when the nervous system is presented with redundant, discrepant beeps and flashes it shifts the timing of the beep if it can infer with reasonable likelihood, that the beep and flash arose from a single event. This response strategy is in line with a unisensory adjustment process (see Di Luca, et al., 2009).

The final study presented in Chapter 7 explored how the nervous system estimates the duration of a redundant sound and flash event. Observers were presented with either two intervals of sound, two intervals defined by flashes, or two intervals of simultaneously presented sounds and flashes. The observers had to indicate if the first or second interval (irrespective of the type of signal, sound, light flash or both) was longer. When they judged the combined sound and flash intervals their estimates were more precise than either the light flash or sound alone. That is, when the nervous system was provided with redundant cues it had an advantage that reduced uncertainty in its interactions with the environment. This finding highlights the benefit observed for multisensory integration.

The findings presented in this thesis cover a range of perceptual challenges faced by the nervous system in our multisensory world. Figure 7.1 outlines an initial suggestion of the mechanisms hypothesised to underlie these processes for both redundant and complementary cues. In the following sections the results summarised above will be discussed in terms of cue properties and potential mechanisms.



Multiple sensory cue integration

Problem faced	Channel 1	Channel 2	Redundant	Complementary	Mechanism
Unisensory or multisensory event?	time	time	Specify different event properties	Complementary given task. Cues compete for representation	Biased competition of complementary cues - possibly occuring in decision-making areas
Discrepant: realign and integrate or segregate?	time	time	Specify same feature but provide discrepant info	Redundant during exposure period	Unisensory adjustment of processing time with gain mechanism - SC & further cortical areas
Duration of one or two events?	time	time	Both cues specify the same feature: event duration	Not complementary	Weighted averaging of redundant cues in polysensory areas with normalisation model
	PERCEPT				

Figure 8.1. The hypothesised mechanisms (right most panel) for the three different cue integration problems investigated in the current thesis. The mechanisms for redundant and complementary cues are proposed to be distinct given the evidence from behavioural outcomes observed in the current tasks.

8.2 Models and mechanisms

8.2.1 Unpredictable sequences and complementary cues

The Colavita effect has been observed for audiovisual and now for visuotactile sensory pairings (Hartcher-O'Brien et al., 2008; Hartcher-O'Brien, Levitan, & Spence, 2010; Hecht & Reiner, 2009; Occelli, Hartcher-O'Brien, Spence, & Zampini, 2010, see Spence, et al., 2012 for a review). This unique phenomenon demonstrates how the nervous system promotes the representation of the visual cue to the multisensory event at the cost of the other sensory cue component. The following sections detail why the nervous system may inhibit the non-visual (i.e., tactile or auditory) cue and carry the visual cue further to awareness. The biased competition hypothesis is proposed as a possible framework to explain the Colavita effect. Desimone and Duncan (1995) conceived of the hypothesis to account for the competition observed at a neural level with the result that certain sensory information reached awareness while other information was inhibited during cue processing.

8.2.1.1 Possible neural underpinning of the Colavita effect

Multisensory interactions occur at all levels of processing in the cortical hierarchy (e.g., Foxe & Schroeder, 2005). In neural terms, it is possible that feedback from higher visual areas influences the representation of (and consequent actions to) the visual and tactile cues in awareness (van Elk, van Schie, Neggers, Bekkering, 2010). As suggested in Section 1.3, visual information processing monopolises brain areas dedicated to processing auditory and tactile information (Kayser et al. 2007; Sereno, Dale, Reppas, Kwong, Belliveau, Brady, Rosen, & Tootell, 1995). Given this, it is more likely that visual cues will not be inhibited by feedback projections from higher order brain regions (Ghazanfar & Schroeder, 2006). Moreover, with respect to the findings of visual dominance over touch reported in this thesis, a number of studies demonstrate that visual information activates the somatosensory cortex for a number of different tasks and cues (e.g., Beauchamp, 2005; Bruce, Desimone, & Gross, 1981; Cohen & Andersen, 2000; Macaluso, Frith, & Driver, 2002). The opposite direction of influence has not been so widely documented. If the representation of the visual stimulus is increased by activation of the somatosensory areas for visual alone stimuli, it is possible that, under conditions of sensory competition, the nervous system may interpret the somatosensory activation as additional visual, rather than tactile input (Desimone & Duncan, 1995). Furthermore, spatial proximity of the complementary cues was shown to influence the visual dominance over touch (Chapters 2 & 3), indicating that this factor alters the cue competition biased in vision's favour.

The Colavita effect appears to depend on the task required of the observer (Ngo et. al., 2011; Sinnett et al., 2008). As such a plausible site for resolving the competition in representation of the responses to the visual flash and tactile pulse cues, in the Colavita effect, are decision-making areas such as the fronto-parietal network and extrastriate areas. These networks could potentially boost motor responses to visual, as compared to tactile cues and, given the advantage that visual input already receives during sensory level processing this could account for the uni-directional nature of the effect (see Duncan et al., 1997, for a review of the influence of such regions). The findings from this dissertation demonstrating involvement of person space distinctions further supports an hypothesis for 'action preparation' and decision-making regions as these have been demonstrated to account for personal and extrapersonal space action repertoires of observers (Committeri et al., 2007; see Spence et al., 2012, for a review).

While the timing of these possible neural influences cannot be speculated on here given the inconsistent RT results of the Colavita experiments in this thesis, it has been proposed that the timing of neural projections from visual to somatosensory and auditory areas is slower than in the reverse direction, and could therefore account for inhibitory responses affecting touch (Spence et al., 2012).

8.2.1.2 Possible explanations drawing on evolution and optimisation

A second possible explanation as to why vision may dominate over touch in the speeded detection Colavita paradigm comes from the patterns of stimulation present in the natural environment. That is, visual information in the environment is constantly changing and therefore information in the visual channel needs to be monitored and updated – i.e., this information will not be inhibited. As a general rule, tactile stimulation changes more slowly. To highlight this, imagine you are sitting on your chair before your desk. The information from the tactile pressure sensors is stable. In this case the nervous system can habituate to the information – i.e., leading to suppression of touch. Due to a lifetime of experience with visual and tactile stimuli, the nervous system has learnt that tactile input is slower than visual to change (e.g., Conrad, Vitello, & Noppeney, in press).

While these tentative explanations are couched in neural terms, it is important to recognise that the Colavita effect is a small, behavioural response and before networks can be described to underpin it, it would be recommendable that future studies used neuroimaging techniques to highlight whether a neural suppression of the tactile (or auditory) cue occurs.

8.2.1.3 Temporal sequences and response patterns

From the work presented in the chapters focusing on temporal recalibration to audiovisual synchrony (see Chapters 4, 5, & 6), an important finding emerged: the nervous system must establish that the two redundant cues arise from a single event (for a review of unity and time see Spence & Parise, 2010), before integrating and recalibrating. This normally occurs when the correlation between redundant visual and auditory cues becomes apparent – i.e., as a result of prolonged exposure. In this case the nervous system establishes the statistics of the current environment and can adapt its response to account for them (Den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010; Fiebelkorn, Foxe, & Molholm, 2010). If the nervous system is presented with a sequence of unisensory and multisensory trials – as occurred both in the Colavita visual dominance task (Chapters 2 & 3) and in the temporal recalibration experiment with intermittent sampling (Chapter 5) – then it appears that under such circumstances it is not possible to establish the statistics of the environment. If this is not possible, then it is consequently not possible to resolve such conflicting aspects of the environmental input, for example, a temporal discrepancy or sensory competition. Therefore, perhaps both effects, recalibration and the Colavita like inhibition of touch, depend on the short-term predictions that the nervous system derives from its immediate environment. When observers were exposed, for short periods, to the redundant, discrepant audiovisual cues, they failed to recalibrate. Perhaps the Colavita paradigm poses a similar problem for prediction. In the Colavita effect, the nervous system cannot predict what the next stimulus will be. It can only ascertain that the probability of unisensory information is (usually) higher than multisensory (with the exception of Experiment 2.2). This lack of predictability results in a response strategy able to deal with the competition arising from the complementary cues and task: inhibiting the inconsistent multisensory cues producing visual-only responses. In the following sections the influence on mechanisms for temporal sequences and consequent short-term predictions are discussed for temporal recalibration.

8.2.2 Temporal recalibration

8.2.2.1 Dealing with continuously presented audiovisual temporal conflict

Many studies have demonstrated that when the nervous system is exposed to conflicting redundant audiovisual information over a long period of time it recalibrates to this discrepancy (see Chapters 1, 3, & 5). The findings from the current thesis highlight a number of facts about the process of temporal recalibration. Foremost, the fact that exposure to audiovisual asynchrony modulates response latencies to unisensory auditory signals indicates that recalibration changes the speed of processing of one of the sensory systems rather than simply updating the relationship between the visual and auditory stimuli. The observed frequency selective response for simultaneity estimates is also consistent with spatiotopic recalibrational effects (Heron et al., 2012). From these two factors it is possible to conclude that recalibration adjusts the timing of one of the senses (perhaps to account for its prediction of when the second component should occur) and recalibration alters the temporal properties of the event but not the identity or other non-transient properties. One possibility is that the change in sensory processing happens at the unisensory level. This would account for the shift in RTs to auditory stimuli following recalibration observed in Chapter 4. The change, however, is not likely to be peripheral and be purely bottom-up since the aftereffect is present when the stimulated ear changes, and when the cues are presented at different stimulus locations.

Temporal recalibration appears to be a general timing mechanism in that it keeps the stimulus features indicative of identity etc. stable, while updating the timing of things in a world in which small temporal discrepancies are introduced continuously by physical, neural, and cognitive factors (see Spence & Squire, 2003). This stabilizing, realignment mechanism affects all aspects of temporal processing from response latency, to perceptual processes including producing a change in the magnitude of the neural response (Stekelenburg, Sugano, & Vroomen, 2011 for sensorimotor recalibration).

Whatever mechanisms generate the recalibration to simultaneity it is interesting to consider their possible functional roles. The perceived temporal shift, which partially re-establishes what is synchronous and what not, generates a clear boundary between what is experienced as an audiovisual event and what is experienced as random and non-related. Thus, recalibration appears to be the tool that the nervous system has to maintain a flexible, while plausible representation of events in the environment irrespective of small discrepancies present between different (redundant) sensory cues. If multiple versus a single event were assumed to cause the audiovisual stimulation then the failure to realign the signals in time (via the responses latencies to them) would be a more adaptive strategy.

Given that recalibration has been shown to generalise to touch when touch was not presented in the exposure phase, it is important to look more closely at what was affected by recalibration in these instances (Di Luca et al., 2009). Di Luca et al. reported that recalibrating to an audiovisual discrepancy transferred to audio-tactile and visuo-tactile pairings depending on which sensory modality appeared to have recalibrated under their different test conditions. This generalisation is distinctly different from the lack of transfer observed in the current thesis for simultaneity estimates (see Chapter 6). Although realignment for the auditory system occurred and the auditory system may have changed its speed of processing, it still maintains the frequency selectivity that drives all auditory processes from the cochlear to the cortex (Moore, 2003).

8.2.2.2 Possible neural mechanisms

Psychophysical data regarding the mechanisms of temporal recalibration is still scarce, and therefore a strong conclusion cannot yet be drawn. However, recordings of SC activation after asynchronous exposure offer some indications of how perceptual changes could be obtained (Yu et al., 2009). Recordings made before and after repeated exposure to redundant asynchronous audiovisual stimuli, indicate a change in the neural responses to the two stimuli. The change operates so as to make them overlap in time despite their asynchrony. The neural response to the second stimulus in the SC has a shorter latency, and the response to the first stimulus has longer duration and overall higher response magnitude. These changes arise gradually as the stimulus pair is repeatedly presented, and they last beyond exposure when stimuli are presented individually. How these changes in neural activity map to perception remains to be understood.

To summarise, the results of several studies indicate that the mechanism involved in such temporal realignment works by gradually adjusting unisensory processing to change perceived timing of the stimuli. This adjustment is obtained starting with early sensory areas in the brain (i.e., the superior colliculus) where adaptation changes the neural responses to the stimuli, making them more coincident in time (see Yu et al., 2009). The nervous system maintains its ability to plastically respond to changes in the sensory inputs in both the short and long term especially for temporal features. However, this plasticity of response has been one element suggested for why the nervous system's integration strategy for redundant temporal cues has been harder to capture than for that of spatial aspects of the environment (see Hartcher-O'Brien & Alais, 2011).

8.2.3 Multisensory time perception and integration: Common mechanisms to spatial cue integration?

Previous studies (e.g., Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; van Beers, Sitting, Denier, & van der Gon, 1999) have demonstrated that redundant *spatial* cues, from multiple sensory channels are averaged according to the reliability of the each sensory systems information. This weighted averaging is optimal in that it gives rise to a multisensory estimate with minimal uncertainty. These findings contrast with other studies that claimed suboptimal integration of multisensory *temporal* estimates, particularly temporal duration (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011; Roach et al., 2006). Although recent research has struggled to identify the most appropriate models of redundant cue sensory integration in the time domain (Burr et al., 2009a; Hartcher-O'Brien & Alais, 2011; Ley et al., 2009; Roach et al., 2006; Shams et al., 2002; Welch & Warren, 1980), the results on multisensory duration provided in

Chapter 7 coupled with Ley et al.'s findings, provide strong evidence that integration in time and space are supported by the same mechanisms or model. That is, sensory integration of redundant cues is the result of an optimal weighting process (e.g., Alais & Burr, 2004; Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Jacobs, 1999; van Beers et al., 1999), whether they specify temporal or spatial features of the environment.

It has been argued that integrating multiple cues into a coherent duration estimate (despite small conflicts in physical duration) is a particularly interesting process because of its post-hoc nature (Hartcher-O'Brien, & Alais, 2011; Hartcher-O'Brien, Di Luca, & Ernst, submitted). Duration, unlike estimates of temporal order, presents a particular challenge to the nervous system as duration itself can only be accurately estimated once the information is no longer available to the senses. Given this, integration of audiovisual duration information is not as straightforward as size or orientation of location estimates. Being optimal in terms of integrating multisensory event duration therefore means that the integration mechanism operates on the stored representation of event duration and its associated variance.

8.2.3.1 Timing mechanisms and clocks

Mechanisms of interval timing, i.e., duration estimation, have focused on the existence of internal clocks which monitor the different sensory systems (Alais & Cass, 2010; Buhusi & Meck, 2009; Kristofferson, 1984; van Rijn & Taatgen, 2008; van Wassenhove, et al., 2008). For MLE to successfully describe the integration process, an assumption of independent channels needs to be satisfied (Heron et al., 2012b). Given this, optimal integration of event duration provides a hint that duration estimation may be obtained through an independent channels mechanism (i.e., as suggested by Heron et al., 2012b).

8.2.4 Redundant and complementary cues

The framework presented in this thesis suggests that the nature of the cues provided to the nervous system modulates what is possible in terms of multisensory interactions. Redundant cues to environmental properties, such as location, or event timing, can be integrated according to a weighted averaging strategy. Complementary cues, on the other hand, are more likely to be in competition and therefore not averaged and integrated. The sensory processing for the multiple sensory cues occurs throughout the nervous system irrespective of cue type, yet the cues fate is determined by its nature.

8.3 Experimental recommendations for future research

The research presented in this thesis provides valuable insight into multisensory integration for complementary and redundant cues. However, in future investigations it would be exciting to explore explicit temporal sequence influences on the complementary cue competition reported during the Colavita effect. If both the Colavita task (Chapters 2 & 3) and the exposure phase from temporal recalibration (Chapters 4 & 6) were combined then it might be possible to see whether the statistics of the short-term environment shape the Colavita effect. For example, if observers were first adapted to a discrepant visuotactile cue pairing, and then presented with the typical unpredictable sequence of unisensory and non-discrepant multisensory cues from the Colavita paradigm, it is hypothesised that the inhibition of touch would not occur: Due to recalibration, an observer's perceived timing of the multisensory cues would be altered by exposure to discrepant visuotactile cues, and the visual and tactile cues in the multisensory test pair. Even though these are presented simultaneously, they should be perceived as non-simultaneous and it is therefore less likely that the tactile cue would be inhibited. However, the result remains for empirical testing.

Furthermore, the null results reported in Chapter 5 present a further opportunity to probe how the discrepancy in short-term exposure to conflicting but redundant cues is resolved. This was discussed in detail in Section 5.4, but one way to overcome the null result observed in this thesis would be to alter the measurement method from RTs to TOJs. The additional noise from the response distribution would then be eliminated and it might be possible to observe a change with short periods of exposure.

While a Kalman filter is one plausible level of explanation for temporal recalibration, the neural level could also provide valuable insights into how this process occurs. To address how the nervous system achieves the realignment of the signals in time, the author of this thesis is currently conducting a study exploring the effects of temporal recalibration on ERP and EEG responses during and after recalibration.

Recent studies (e.g., Heron et al., 2012a; Navarra et al., 2012) have attempted to understand the influences of reference frame on temporal recalibration. In relation to the observed frequency specific recalibration effect in Chapter 6, the question that emerged was whether recalibration is a selective reference frame effect or not, assuming auditory space is coded tonotopically. One proposed experiment would involve monaural temporal recalibration, which would encompass both the tonotopic and spatial co-ordinate components of audiovisual temporal recalibration. If recalibration occurred and did not transfer across ears, then it should be one way of realigning the auditory maps (spatial via temporal). This experiment is planned for the author's post-doctoral work.

8.4 Concluding remarks: Implications for multisensory integration

Let us return, for a final time, to the example of locating the person calling your name. When cues to this event, given your task, are complementary, their role is generally to increase the richness of the representation formed of the environmental event as a whole. For example the familiar smell of perfume will not help you localise your friend but will increase your certainty that the individual is known to you. Imagine you are scanning the crowd to locate your friend. Your ability to suppress the non-visual information arising from the perfume or the chatter of conversations around you is similar to what is observed in the Colavita effect. The task your nervous system is solving can either set up competition between the multisensory cues for representation in awareness, or if both cues are redundantly specifying the location then the multisensory nature of the input allows your nervous system to average across the two sources and optimise your estimate of where and who. In generalizing to other non-measured multisensory scenarios, the current thesis provides strong evidence for the fact that the nervous system selects its response flexibly depending on the nature of the cues and the environmental constraints. Given these, it will optimise its processing of the input and produce the most appropriate response.

References

Abrahms, R., Davoli, F., Du, C., Knapp, W., & Paull, D. (2008). Altered vision near the hands. *Cognition*, **107**, 1035–1047.

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

Alais, D., & Carlile, S. (2005). Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound. *Proceedings of the National Academy of Sciences, USA*, **102**, 2244–2267.

Alais, D., & Cass, J. (2010). Multisensory perceptual learning of temporal order: Audiovisual learning transfers to vision but not audition. *PLoS ONE*, **5**, e11283.

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, **23**, 3–38.

Allan, L. G., & Kristofferson, A. B. (1974). Psychophysical theories of duration discrimination. *Perception & Psychophysics*, **16**, 26–34.

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

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology and computation. *Current Opinion in Neurobiology*, **19**, 452–458.

Arnold, D., & Yarrow, K. (2011). Temporal recalibration of vision. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 535–538.

Bach-y-Rita, P. (1967). Sensory plasticity: Applications to a vision substitution system. *Acta Neurologica*, **43**, 417–426.

Baddeley, R. J., Ingram, H. A., & Miall, R. C. (2003). System identification applied to a visuomotor task: Near-optimal human performance in a noisy changing task. *Journal of Neuroscience*, **23**, 3066–3075.

Banich, M. T. (1998). The missing link: The role of interhemispheric interaction in attentional processing. *Brain and Cognition*, **36**, 128–157.

Balci, F., Freestone, D., Simen, P., Desouza, L., Cohen, J. D., Holmes, P. (2011). Optimal temporal risk assessment. *Frontiers in Integrative Neuroscience*, **5**, doi: 10.3389/fnint.2011.00056.

Bao, M., Yang, L., Rios, C., He, B., & Engel, S. A. (2010). Perceptual learning increases the strength of the earliest signals in the visual cortex. *Journal of Neuroscience*, **30**, 15080–15084.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, **20**, 1391–1397.

Beauchamp, M. S., (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, **3**, 93–114.

Bedford, F. L. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 232–248.

Bedford, F. L. (2001). Towards a general law of numerical/object identity. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition*, **20**, 113–175.

Behrmann, M., Moscovitch, M., & Mozer, M. C. (1991). Directing attention to words and nonwords in normal subjects and in a computational model: Implications for neglect dyslexia. *Cognitive Neuropsychology*, **8**, 213–248.

Bender, M. B. (1952). *Disorders of perception.* Springfield, IL: Charles C. Thomas.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, **13**, 90–102.

Bertelson, P. (1998). Starting from the ventriloquist: the perception of multimodal events. In M. Sabourin (Eds.), *Advances in psychological science* (pp. 419–439). Hillsdale, NJ: Erlbaum.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, **5**, 482–489.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension 1. Evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology*, **50**, 147–155.

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–178). Oxford, UK: Oxford University Press.

Borgers, S., Hauser, T.-K., & Himmelbach, M. (2011). Bilateral hand representations in human primary proprioceptive areas. *Neuropsychologia*, **49**, 3383–3391.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, **391**, 756.

Brayanov, J. B., & Smith, M. A. (2010). Bayesian and "Anti-Bayesian" biases in sensory integration for action and perception in the size-weight illusion. *Journal of Neurophysiology*, **103**, 1518–1531.

Bremner, A., Lewkowicz, D., & Spence, C. (Eds.) (2012). *Multisensory development*. Oxford, UK: Oxford University Press.

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, **6**, 554–564.

Brewster, D. (1832). *Letters on natural magic, addressed to Sir Walter Scott, Bart* (pp. 11–13). London, UK: John Murray, Albemarle Street.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, **46**, 369–384.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, **6**, 755–765.

Buhusi, C. V., & Meck, W. H. (2009). Relativity theory and time perception: Single or multiple clocks? *PLoS ONE*, **4**, e6268.

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

Burge, J., Ernst, M. O., & Banks, M. S. (2008). The statistical determinants of adaptation rate in human reaching. *Journal of Vision*, **8**, 21–29.

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *Journal of Neuroscience*, **30**, 7714–7721.

Burgoon, J. K., & Saine, T. (1978). *The unspoken dialogue: An introduction to nonverbal communication* (p. 21). Boston, MA: Hughton Mifflin Company.

Burr, D., Banks, M. S., & Morrone, M. C. (2009a). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, **198**, 49–57.

Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., & Morrone, M. C. (2009b). Temporal mechanisms of multimodal binding. *Proceedings of the Royal Society of London, B: Biological Sciences*, **276**, 1761–1769.

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.

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading [Abstract]. *Journal of Vision*, **10**, 23.

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Cardini, F., Longo, M. R., Driver, J., & Haggard, P. (in press). Rapid enhancement of touch from non-informative vision of the hand. *Neuropsychologia*, doi:10.1016/j.neuropsychologia.2012.04.020.

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**, 1–14.

Carpenter, R. H., & Williams, M. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, **377**, 59–62.

Carr, C. E., & Konishi, M. (1990). A circuit for detection of interaural time differences in the brainstem of the barn owl. *Journal of Neuroscience*, **10**, 3227–32246.

Cate, A., & Behrmann, M. (2002). Spatial and temporal influences on extinction. *Neuropsychologia*, **40**, 2206–2225.

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, **32**, 1056–1060.

Cohen, Y. E., & Andersen, R. A. (2000). Reaches to sounds encoded in an eye-centred reference frame. *Neuron*, **27**, 647–652.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, **16**, 409–412.

Colavita, F. B., Tomko, R., & Weisberg, D. (1976). Visual prepotency and eye orientation. *Bulletin of the Psychonomic Society*, **8**, 25–26.

Colavita, F. B., & Weisberg, D. (1979). A further investigation of visual dominance. *Perception & Psychophysics*, **25**, 345–347.

Committeri, G., Pitzalis, S., Galati, G., Patria, F., Pelle, G., Sabatini, U., Castriota-Scanderbeg, A., Piccardi, L., Guariglia, C., & Pizzamiglio, L. (2007). Neural bases of personal and extrapersonal neglect in humans. *Brain*, **130**, 431–441.

Congedo, M., Lécuyer, A., & Gentaz, E. (2006). The influence of spatial delocation on perceptual integration of vision and touch. *Presence: Teleoperators & Virtual Environments*, **15**, 353–357.

Conrad, V., Vitello, M. P., & Noppeney, U. (in press). Interactions between apparent motion rivalry in vision and touch. *Psychological Science*.

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

Costantini, M., Bueti, D., Pazzaglia, M., & Aglioti, S. M. (2007). Temporal dynamics of visuo-tactile extinction between and within-hemispaces. *Neuropsychology*, **21**, 242–250.

Clarke, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Norwell: Kluwer Academic Publishers.

Cynx, J., & Shapiro, M. (1986). Perception of missing fundamental by a species of songbird (Sturnus Vulgaris). *Journal of Computational Psychology*, **100**, 356–360.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, **7**, 460–467.

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology – Paris*, **98**, 249–258.

den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, **30**, 3210–3219.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA*, **93**, 13494–13499.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, **18**, 193–222.

Di Luca, M., Machulla, T., & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, **9**, 1–16.

di Pellegrino, G., Làdavas, E., & Farné, A. (1997). Seeing where your hands are. *Nature*, **388,** 730.

Dixon, N., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719–721.

Dufour, A., & Touzalin, P. (2008). Improved sensitivity in the perihand space. *Experimental Brain Research*, **190**, 91–98.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, **7**, 255–261.

Eagleman, D. M., Tse, P. U., Buonomano, D., Janssen, P., Nobre, A. C., & Holcombe, A. (2005). Time and the brain: how subjective time relates to neural time. *Journal of Neuroscience*, **25**, 10369–10371.

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

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science*, **305**, 875–877.

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

Epstein, W. (1975). Recalibration by pairing: A process of perceptual learning. *Perception*, **4**, 59–72.

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., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, **8**, 162–169.

Ernst, M. O., & Di Luca, M. (2011). Multisensory perception: from integration to remapping. In J. Trommershäuser, M. S. Landy, & K. Körding (Eds.), *Sensory cue integration* (pp. 224–250). New York, NY: Oxford University Press.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 1058–1079.

Fain, G. (2003). *Sensory transduction*. Sunderland, MA: Sinauer.

Farné, A., & Ladavas, E. (2000). Dynamic size-change of hand peripersonal space following tool use. *NeuroReport*, **11**, 1645–1649.

Farné, A., Pavani, F., Meneghello, F., & Làdavas, E. (2000). Left tactile extinction following visual stimulation of a rubber hand. *Brain*, **123**, 2350–2360.

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, **63**, 719–725.

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: How much do learned associations matter? *Cerebral Cortex*, **20**, 109–120.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, **1,** 3–32.

Foulkes, A., & Miall, R. (2000). Adaptation to visual feedback delays in a human manual tracking task. *Experimental Brain Research*, **131**, 101–110.

Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, **16**, 419–423.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, **147**, 332–343.

Frassinetti, F., Maini, M., Benassi, M., Avanzi, S., Cantagalloc, A., & Farnè, A. (2010). Selective impairment of self body-parts processing in right brain-damaged patients. *Cortex*, **46**, 322–328.

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

Gallace, A., Auvray, M., Tan, H. Z., & Spence, C. (2006). When visual transients impair tactile change detection: A novel case of crossmodal change blindness? *Neuroscience Letters*, **398**, 280–285.

Gallace, A., & Spence, C. (2005). Visual capture of apparent limb position influences tactile temporal order judgments. *Neuroscience Letters*, **379**, 63–68.

Gallace, A., & Spence, C. (2007). The cognitive and neural correlates of "tactile consciousness": A multisensory perspective. *Consciousness and Cognition*, **17**, 370–407.

Gallace, A., Tan, H. Z., & Spence, C. (2007). The body surface as a communication system: The state of the art after 50 years. *Presence*: *Teleoperators & Visual Environments*, **16**, 655–676.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, **116**, 439–453.

Gamache, P.-L., & Grondin, S. (2010). Sensory-specific clock components and memory mechanisms: Investigation with parallel timing. *European Journal of Neuroscience*, **31**, 1908–1914.

Gephstein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, **5**, 1013–1023.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, **10**, 278–285.

Giard, M. H., & Peronnet, F. (1999). Auditory–visual integration during multimodal object recognition in humans: A behavioural and electrophysiological study. *Journal of Cognitive Neuroscience*, **11**, 473–490.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, **84**, 279–325.

Gibbon, J., & Church, R. M. (1984). Sources of variance in an information processing theory of timing. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 465–488). Hillsdale, NJ: Erlbaum.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. In J. Gibbon & L. G. Allan (Eds.), *Timing and time perception* (**423**, pp. 52–77). New York, NY: The New York Academy of Sciences.

Gibson, J. J. (1943). Adaptation, after-effect and contrast in the perception of curved lines. *Journal of Experimental Psychology*, **16**, 1–31.

Graighero, L., Fadiga, L., Umilta, C., & Rizzolati, G. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 1673–1692.

Graziano, M. S., & Gross, C. G. (1994). The representation of extrapersonal space: a possible role of bimodal visual-tactile neurons. In M. S. Grazzaniga (Ed.), *The cognitive neurosciences* (pp. 1021–1034). Cambridge, MA: MIT Press.

Graziano, M. S., & Gross, C. G. (1998). Spatial maps for the control of movement. *Current Opinion in Neurobiology*, **8**, 195–201.

Greenwald, H. S., Knill, D., & Saunders, J. A. (2005). Integrating visual cues for motor control. *Vision Research*, **45**, 1975–1989.

Gregory, R. L. (1967). Origin of eyes and brains. *Nature*, **213**, 369–372.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, **54**, 383–394.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, **127**, 22–44.

Grondin, S., Ouellet, B., & Roussel, M.-È. (2001). About optimal timing and stability of Weber fraction for duration discrimination. *Acoustical Science & Technology*, **22**, 370–372.

Guest, S., & Spence, C. (2003). Tactile dominance in the speeded discrimination of textures. *Experimental Brain Research*, **150**, 201–207.

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008a). Asynchrony adaptation is not specific to sound and vision. *Ophthalmic and Physiological Optics*, **28**, 96–97.

Hanson, J. V. M., Heron, J., & Whitaker, D. (2008b). Recalibration of perceived time across sensory modalities. *Experimental Brain Research*, **185**, 347–352.

Hanson, J. V. M., Whitaker, D., & Heron, J. (2009). Preferential processing of tactile events under conditions of divided attention: effects of divided attention on reaction time. *Neuroreport*, **7,** 1392-1396.

Hari, R., & Jousmäki, V. (1996). Preference of personal to extrapersonal space in a visuomotor task. *Journal of Cognitive Neuroscience*, **8**, 305–307.

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.

Hartcher-O'Brien, J., & Alais, D. (2011). Temporal ventriloquism in a purely temporal context. *Journal of Experimental Psychology: Human Perception & Performance*, **37**, 1383–1395.

Hartcher-O'Brien, J., Di Luca, M., & Ernst, M. O. (submitted). The duration of uncertain times: Audiovisual information about intervals is integrated in a statistically optimal fashion.

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.

Hartcher-O'Brien, J., Levitan, C. A., & Spence, C. (2010). Extending the visual dominance of touch for input off the body. *Brain Research*, **1362**, 48–55.

Hawkins, D. L. (1989). Using U statistics to derive the asymptotic distribution of Fisher's Z statistic. *The American Statistician*, **43,** 235–237.

Hay, J. C., Pick, H. L., Jr., & Ikeda, K. (1965). Visual capture produced by prism spectacles. *Psychonomic Science*, **2**, 215–216.

Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, **193**, 307–314.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, **179**, 595–606.

Helbig, H. B., & Ernst, M. O. (2008). Haptic perception in interaction with other senses. *Human Haptic Perception: Basics and Applications, III*, 235–249.

Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2012b). Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 690–698.

Heron, J., Roach, N., Hanson, J. V. M, McGraw, P., & Whitaker, D. (2012a). Audiovisual time perception is spatially specific. *Experimental Brain Research*, **218**, 477–485.

Heron, J., Roach, N., Whitaker, D., & Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing. *European Journal of Neuroscience*, **31**, 1755–1762.

Heron, J., Whitaker, D., McGraw, P., & Horoshenkov, K. (2007). Adaptation minimizes distance-related audiovisual delays. *Journal of Vision*, **7**, 1–8.

Heymans, G. (1899). Untersuchungen über psychische Hemmung [Investigations of psychological inhibition]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, **21**, 321–359.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, **4**, 967–992.

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

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors*, **49**, 1107–1114.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, **11**, 157–174.

Ho, C., Tan, H. Z., & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, **8**, 397–412.

Holmes, N., Snijders, D., & Spence, C. (2006). Reaching with alien limbs: Visual exposure to prosthetic hands biases proprioception without accompanying illusions of ownership. *Perception & Psychophysics*, **68**, 685–701.

Holmes, N. P., & Spence, C. (2006). Beyond the body schema: Visual, prosthetic, and technological contributions to bodily perception and awareness. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 15–64). Oxford, UK: Oxford University Press.

Holmes, N. P., & Spence, C. (2007). Dissociating body image and body schema with rubber hands. *Behavioral & Brain Sciences*, **30**, 211–212.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London, UK and New York, NY: John Wiley and Sons.

Igarashi, Y., Kitagawa, N., Spence, C., & Ichihara, S. (2007). Assessing the influence of schematic drawings of body parts on tactile discrimination performance using the crossmodal congruency task. *Acta Psychologica*, **124**, 190–208.

Inan, S., Mitchell, T., Song, A., Bizzell, J., & Belger, A. (2004). Hemodynamic correlates of stimulus repetition in the visual and auditory cortices: an fMRI study. *Neuroimage*, **21**, 886–893.

Inhoff, K. W., Rafal, R. D., & Posner, M. I. (1992). Bimodal extinction without crossmodal extinction. *Journal of Neurology, Neurosurgery and Psychiatry*, **45**, 147–161.

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, **12**, 273–280.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, **39**, 3621–3629.

Jacobs, S., Brozolli, C., & Farné, A. (2012). Neglect: A multisensory deficit. *Neuropsychologia*, **50**, 1029–1044.

Jacobson, E. (1911). Experiments on the inhibition of sensations. *Psychological Review*, **18**, 24–53.

Jaskowski, P. (1996). Simple reaction time and perception of temporal order: Dissociations and hypotheses. *Perceptual & Motor Skills*, **82**, 707–730.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, **13**, 1020–1026.

Jeffress, L. (1948). A place theory of sound localization. *Journal of Comparative Physiology and Psychology*, **41**, 35–39.

Johnston, A., Arnold, D., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology*, **16**, 472–479.

Kagerer, F. A., Contreras-Vidal, J. L., & Stelmach, G. E. (1997). Adaptation to gradual as compared with sudden visuo-motor distortions. *Experimental Brain Research*, **115**, 557–561.

Kammers, M., de Vignemont, F., Verhagen, L., & Dijkerman, H. C. (2009). The rubber hand illusion in action. *Neuropsychologia*, **47**, 204–211.

Kao, K. L., & Goodale, M. A. (2008). Enhanced detection of visual stimuli projected on a tool [Abstract]. *Journal of Vision*, **8**, 1061.

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, **53**, 427–438.

Kayser, C., & Logothetis, N. K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Structure & Function*, **212**, 121–132.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience*, **27**, 1824–1835.

Keetels, M. N., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, **167**, 635–640.

Keetels, M. N., & Vroomen, J. (2010). Perception of synchrony between the senses. In M. M. Murray & M. T. Wallace (Eds.), *Frontiers in the neural basis of multisensory processes* (pp. 1–61). London, UK: Taylor & Francis.

King, A. J. (2004). The superior colliculus. *Current Biology*, **14**, R335-8.

King, A. J. (2005). Multisensory Integration: Strategies for synchronization. *Current Biology*, **15**, R339-R341.

King, A. J., & Calvert, G. (2001). Multisensory integration: Perceptual grouping by eye and ear. *Current Biology*, **11**, R322-R325.

King, A. J., & Palmer, A. R. (1983). Cells responsive to free-field auditory stimuli in guinea-pig superior colliculus: Distribution and response properties. *The Journal of Physiology*, **342**, 361–381.

King, A. J., Parsons, C. H., & Moore, D. R. (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proceedings of the National Academy of Sciences, USA*, **97**, 11821–11828.

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature*, **416**, 172–174.

Klein, R. M. (1977). Attention and visual dominance: A chronometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 365–378.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, **43**, 2539–2558.

Knudsen, E. I., Knudsen, P. F., & Esterly, S. D. (1984). A critical period for the recovery of sound localization accuracy following monaural occlusion in the barn owl. *The Journal of Neuroscience*, **4**, 1012–1020.

Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. *Science*, **200**, 795–797.

Kopinska, A., & Harris, L. R. (2004). Simultaneity constancy. *Perception*, **33**, 1049–1060.

Koppen, C., Alsius, A., & Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect. *Experimental Brain Research*, **184**, 533–546.

Koppen, C., Levitan, C., & Spence, C. (2009). A signal detection study of the Colavita effect. *Experimental Brain Research*, **196**, 353–360.

Koppen, C., & Spence (2007a). Seeing the light: Exploring the Colavita visual dominance effect. *Experimental Brain Research*, **180**, 737–754.

Koppen, C., & Spence, C. (2007b). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters*, **417**, 107–111.

Koppen, C., & Spence, C. (2007c). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters*, **418**, 266–271.

Koppen, C., & Spence, C. (2007d). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research*, **1186**, 224–232.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, **2**, e943.

Kristofferson, A. B. (1984). Quantal and deterministic timing in human duration discrimination. *Annals of the New York Academy of Sciences*, **423**, 3–15.

Kuschel, M., Di Luca, M., Buss, M., & Klatzky, R. L. (2010). Combination and integration in the perception of visual-haptic compliance information. *IEEE Transactions on Haptics*, **3**, 234–244.

Ladavas, E., di Pellegrino, G., & Zeloni, G. (1998). Neuropsychological evidence of an integrated visuotactile representation of peripersonal space in humans. *Journal of Cognitive Neuroscience*, **10**, 581–589.

Ladavas, E., Farnè, A., Zeloni, G., & di Pellegrino, G. (2000). Seeing or not seeing where your hands are. *Experimental Brain Research*, **131**, 458–467.

Lalanne, C., & Lorenceau, J. (2004). Crossmodal integration for perception and action. *Journal of Physiology – Paris*, **98**, 265–279.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, **35**, 389–412.

Lane, A. R., Ball, K., Smith, D. T., Schenk, T., & Ellison, A. (2011). Near and far space: Understanding the neural mechanisms of spatial attention. *Human Brain Mapping,* doi:10.1002/hbm.21433.

Lederman, S. J., Thorne, G., & Jones, B. (1986). Perception of texture by vision and touch: Multidimensionality and intersensory integration. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 169–180.

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research*, **121**, 69–79.

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, P. A., & Miall, R. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia*, **41**, 1583–1592.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 1094–1106.

Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and vibrotactile information for judgments of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, **35**, 1005–1019.

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, **1173**, 102–109.

Lockhart, D. & Ting, L. (2007). Optimal sensorimotor transformations for balance. *Nature Neuroscience*, **10**, 1329-1336.

Ma, W. J., Beck, J. M., & Pouget, A. (2008). Spiking networks for Bayesian inference and choice. *Current Opinion in Neurobiology*, **18**, 217–222.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, **289**, 1206–1208.

Machulla, T.-K., Di Luca, M., & Ernst, M. O. (2010). Does audiovisual temporal recalibration store without stimulation? [Abstract]. *Journal of Vision*, **10**, 1414.

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

Makin, T., Holmes, N. P., & Ehrsson, H. H. (2008). On the other hand: Dummy hands and peripersonal space. *Behavioural Brain Research*, **191**, 1–10.

Makin, T., Holmes, N. P., & Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in the human intraparietal sulcus. *Journal of Neuroscience*, **27**, 731–740.

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: Close to hand and within reach. *Current Biology*, **13**, R531-R539.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: W. H. Freeman and Company.

Matell, M. S., & Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays*, **22**, 94–103.

Mattingley, J. B., Driver, J., Beschin, N., & Robertson, I. H. (1997). Attentional competition between modalities: Extinction between touch and vision after right hemisphere damage. *Neuropsychologia*, **35**, 867–880.

Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review Neuroscience*, **27**, 307–340.

McAlpine, D., Jiang, D., & Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, **4**, 396–401.

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

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *Journal of the Acoustical Society of America*, **77**, 678–685.

Meredith, M. A. (2004). Corticocortical connectivity of cross-modal circuits. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 343–356). Cambridge, MA: MIT Press.

Meredith, M. A., Nemitz, J., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *Journal of Neuroscience*, **7**, 3215–3229.

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.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, **27**, 4120–4131.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, **9**, 875–877.

Moiseff, A. (1989). Binaural disparity cues available to the barn owl for sound localization. *Journal of Comparative Physiology A*, **164**, 629–636.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th Ed.). Bingley, UK: Emerald.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, **17**, 154–163.

Navarra, J., García-Morera, J., & Spence, C. (2012). Temporal adaptation to audiovisual asynchrony generalizes across different sound frequencies. *Frontiers in Psychology*, **3**, 1–7.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences, USA*, **106**, 9169–9173.

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.

Ngo, M. K., Cadieux, M. L., Sinnett, S., Soto-Faraco, S., & Spence, C. (2011). Reversing the Colavita visual dominance effect. *Experimental Brain Research*, **214**, 607–618.

Nico, D. (1999). Effectiveness of sensory stimulation on tactile extinction. *Experimental Brain Research*, **127**, 75–82.

Nityananda, V., & Bee, M. A. (2011). Finding your mate at a cocktail party: Frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PLoS ONE*, **6**, e21191.

Occelli, V., Hartcher-O'Brien, J., Spence, C., & Zampini, M. (2010). Assessing the audiotactile Colavita effect in near and rear space. *Experimental Brain Res*earch, **203**, 517–532.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, **43**, 2451–2468.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, **42**, 74–81.

Pantev, C., Hoke, M., Lutkenhoner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science*, **27**, 486–488.

Parise, C. V., Ernst, M. O., & Spence, C. (2011). When correlation implies causation in multisensory integration. *Current Biology*, **22**, 46–49.

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, **283**, 1272–1273.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, **116**, 220–244.

Pashler, H., & Baylis, G. (1991). Procedural learning: II. Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and cognition*, **17**, 33–48.

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological Science*, **11**, 353–359.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgements of spatial direction. *Perception & Psychophysics*, **6**, 203–205.

Pollux, P. M. J., & Bourke, P. A. (2008). Influence of hand position on the near effect in 3-D attention. *Perception & Psychophysics*, **70**, 1383–1392.

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences,* **1**, 56–61.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, **83**, 157–171.

Quinlan, P. (2000). The 'late' locus of visual dominance. *Abstracts of the Psychonomic Society*, **5**, 64.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *The Quarterly Journal of Experimental Psychology*, **26**, 63–71.

Randich, A., Klein, R. M., & LoLordo, V. M. (1978). Visual dominance in the pigeon. *Journal of the Experimental Analysis of Behavior*, **30**, 129–137.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences, USA*, **95**, 869–875.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, **89**, 1078–1093.

Reed, C., Grubb, J., & Steele, C. (2006). Hands up: Attentional prioritisation of space near the hands. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 166–177.

Rock, I., & Harris, C. S. (1967). Vision and touch. *Scientific American*, **216** (5), 96–104.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, **143**, 594–596.

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.

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2010). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1314–1322.

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.

Roseboom, W., & Arnold, D. H. (2011). Twice upon a time: Multiple concurrent temporal recalibrations of audiovisual speech. *Psychological Science*, **22**, 872–877.

Santangelo, V., & Spence, C. (2007). Assessing the automaticity of reflexive tactile attentional orienting. *Perception*, **36**, 1497–1505.

Sarri, M., Blankenburg, F., & Driver, J. (2006). Neural correlates of crossmodal visual-tactile extinction and of tactile awareness revealed by fMRI in a right-hemisphere stroke patient. *Neuropsychologia*, **44**, 2398–2410.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, **12**, 101–108.

Sekuler, R., Sekuler, A., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., Rosen, B., & Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, **268**, 889–893.

Shams, L. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, **7**, 269–284.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear: Sound induced visual flashing. *Nature*, **408**, 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, **14**, 147–152.

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, **12**, 3849–3852.

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

Shapiro, K. L., & Johnson, T. L. (1987). Effects of arousal on attention to central and peripheral visual stimuli. *Acta Psychologica*, **66**, 157–172.

Simonotto, E., Riani, M., Seife, C., Roberts, M., Twitty, J., & Moss, F. (1997). Visual perception of stochastic resonance. *Physical Review Letters*, **78**, 1186–1189.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, **28**, 1059–1074.

Sinnett, S., Soto-Faraco, S., & Spence, C. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica*, **128**, 153–161.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics*, **69**, 673–686.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science & Technology*, **28**, 61–70.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research*, **176**, 245–258.

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

Spence, C., & Driver, J. (Eds.) (2004). *Crossmodal space and crossmodal attention*. Oxford, UK: Oxford University Press.

Spence, C., & Gallace, A. (2008). Making sense of touch: A multisensory approach to the perception of objects. In E. Pye (Ed.), *Magic touch* (pp. 21–40). London, UK: UCL Press.

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., & Parise, C. V. (2010). Prior entry: A review. *Consciousness & Cognition*, **19**, 364–379.

Spence, C., Parise, C., & Chen, Y.-C. (2012). The Colavita visual dominance effect. In M. M. Murray & M. Wallace (Eds.), *Frontiers in the neural bases of multisensory processes* (pp. 523–550). Boca Raton, FL: CRC Press.

Spence, C., Pavani, F., Maravita, A., & Holmes, N. P. (2008). Multi-sensory interactions. In M. C. Lin & M. A. Otaduy (Eds.), *Haptic rendering: Foundations, algorithms, and applications* (pp. 21–52). Wellesley, MA: AK Peters.

Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, **130**, 799–832.

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology*, **13**, R519-R521.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, **8**, 497–506.

Stein, B. E., & Meredith, M. (1993). *The merging of the senses* (pp. 1–25). Cambridge, MA: MIT Press.

Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, W. J., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 25–50). Oxford, UK: Oxford University Press.

Stekelenburg, J. J., & Vroomen, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism. *NeuroReport*, **16**, 641–644.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, **19**, 1964–1973.

Stekelenburg, J. J., Sugano, Y., & Vroomen, J. (2011). Neural correlates of motor-sensory temporal recalibration. *Brain Research*, **1397**, 46–54.

Sternberg, S., & Knoll, R. (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, NY: Academic Press.

Stetson, C., Cui, X., Montague, P., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, **51**, 651–659.

Stratton, G. M. (1899). The spatial harmony of touch and sight. *Mind*, **8**, 492–550.

Sugano, Y., Keetels, M., & Vroomen, J. (2010). Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Experimental Brain Research*, **201**, 393–399.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, **91**, 1282–1296.

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the 'internal clock'. *Psychological Monographs: General and Applied*, **77**, 1–31.

Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research*, **155**, 301–310.

Uetake, K., & Kudo, Y. (1994). Visual dominance over hearing in feed acquisition procedure of cattle. *Applied Animal Behaviour Science*, **42**, 1–9.

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Perceived duration of expected and unexpected stimuli. *Psychological Research*, **70**, 77–87.

Valenza, N., Murray, M. M., Ptak, R., & Viulleumier, P. (2004). The space of senses: impaired crossmodal interactions in a patient with Balint syndrome after bilateral parietal damage. *Neuropsychologia*, **42**, 1737–1748.

Vallar, G., Rusconi, M. L., Bignamini, L., Geminiani, G., & Perani, D. (1994). Anatomical correlates of visual and tactile extinction in humans: A clinical CT scan study. *Journal of Neurology, Neurosurgery, and Psychiatry*, **57**, 464–470.

van Beers, R. J., Sittig, A. C., & Denier van de Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, **81**, 1355–1364.

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

van Elk, M., van Schie, H. T., Neggers, S. F. W., & Bekkering, H. (2010). Neural and temporal dynamics underlying visual selection for action. *Journal of Neurophysiology*, **104**, 972–983.

van Erp, J. B. F., & van Veen, H. A. H. C. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F*, **7**, 247–256.

van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica*, **129**, 365–375.

van Wassenhove, V., Buonomano, D. V., Shimojo, S., & Shams, L. (2008). Distortions of subjective time perception within and across senses. *PLoS ONE*, **3**, e1437.

van Wassenhove, V., Grant, K. W., & Pöppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, USA*, **102**, 1181–1186.

van Wassenhove, V., Grant, K. W., & Pöppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, **45**, 598–607.

Vatakis, A., Ghazanfar, A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, **8**, 1–11.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2007). Temporal recalibration during asynchronous audiovisual speech perception. *Experimental Brain Research*, **181**, 173–181.

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.

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the "unity assumption" using audiovisual speech stimuli. *Perception & Psychophysics*, **69**, 744–756.

Von Haller Gilmer, B. (1960). Possibilities of cutaneous electro-pulse communication. In G. R. Hawkes (Ed.), *Symposium on cutaneous sensitivity* (pp. 76–84). Fort Knox, KY: U.S. Army Medical Research Laboratory.

Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, **30**, 513–518.

Vroomen, J., & Keetels, M. N. (2009). Sounds change four-dot masking. *Acta Psychologica*, **130**, 58–63.

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention Perception & Psychophysics*, **72**, 871-884.

Vroomen, J., Keetels, M. N., 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., Irion, A. L., & Gordon, D.G. (1981) Simple and contingent aftereffects of perceived duration in vision and audition. *Perception & Psychophysics*, **29**, 475–486.

Ward, R., Goodrich, S., & Driver, J. (1994). Grouping reduces visual extinction. *Visual Cognition*, **1**, 101–129.

Wark, B., Lundstrom, B. N., &, Fairhall, A. (2007). Sensory adaptation. *Current Opinion in Neurobiology*, **17**, 423–429.

Warren, D. H., & Rossano, M. J. (1991). Intermodality relations: Vision and touch. In M. A. Heller & W. Schiff (Eds.), *The psychology of touch* (pp. 119–137). London, UK: Lawrence Erlbaum Associates.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition. *Psychological Review*, **104**, 90–122.

Welch, G., & Bishop, G. (1997). An introduction to the Kalman filter. http://clubs.enscachan.fr/krobot/old/data/positionnement/kalman.pdf.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In G. Ashersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam, NL: Elsevier Science B.V.

Welch, R. B., Duttonhurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, **39**, 294–300.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **3**, 638–667.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and performance: Vol. 1. Sensory processes and perception* (pp. 25–36). New York, NY: Wiley.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Attention, Perception, & Psychophysics,* **63**, 1293–1313.

Wiener, M., Matell, M. S., & Coslett, H. B. (2011). Multiple mechanisms for temporal processing. *Frontiers in Integrative Neuroscience*, **5**, 1–3.

Winter, R., Harrar, V., Gozdzik, M., & Harris, L. R. (2008). The relative timing of active and passive touch. *Brain Research*, **1242**, 54–58.

World Medical Association (2000). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association*, **284**, 3043–3045.

Wozny, D., Beierholm, U., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, **8**, 1–11.

Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *Journal of Neuroscience*, **24**, 1617–1626.

Yarrow, K., Roseboom, W., & Arnold, D. H. (2011). Spatial grouping resolves ambiguity to drive temporal recalibration. *Journal of Experimental Psychology: Human Perception & Performance*, **37**, 1657–1661.

Yu, L., Stein, B., & Rowland, B. (2009). Adult plasticity in multisensory neurons: short-term experience-dependent changes in the superior colliculus. *Journal of Neuroscience*, **29**, 15910-15922.

Zaidel, A., Turner, A. H., & Angelaki, D. E. (2011). Multisensory calibration is independent of cue reliability. *Journal of Neuroscience*, **31**, 13949–13962.

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 judgements: The role of hemispheric redundancy. *International Journal of Psychophysiology*, **50**, 165–180.

Acknowledgements

I would like to thank Prof. Charles Spence and Prof. Marc Ernst for their supervision, and for giving me the freedom to do my own thing. I have learned so much from them both. And from the bottom of my heart I am grateful to Marc for believing in my ideas and my ability to realize them in his lab as my own body of work.

I would also like to thank my unofficial supervisors, Dr. Massimiliano Di Luca and Dr. Jordi Navarra for their day-to-day contributions to my DPhil experience. Our discussions, experimenting and our work together have inspired what I will continue to do. Thank you both for teaching me so much. It should also be said that I have not only learnt things about human perception and research from you, but also what it means to be a scientist and to juggle the personal and the work related aspects of this. Grazie domine pere!

I would have never switched from auditory perception to multisensory perception if it had not been for Assoc. Prof. David Alais – thanks Dave!

To all my colleagues at the Max Planck Institute and at the Department of Experimental Psychology I want to thank you all for the inspiring intellectual interactions and the comradery that has been a grounding force in all the location changes that have occurred in this process. And a special thanks to the brave people who read through parts of this thesis and gave me valuable feedback: Verena Conrad, Jan Souman, Regine Armann, Bron Crosby and Ruth Hartcher-O'Brien – thank you, you all deserve a medal.

My DPhil has spanned two countries, two labs and many dear friends and colleagues. One of the wonderful aspects of research is that it offers you the chance to meet and get to know so many amazing people. So thank you, you amazing people: Malika, Alberto, Cesare, Gustavo, Mark, Carmel, Josh, Verena, Lewis, Tonja, John, Mirko, Loes, Andreas, Marieke, Myrthe, Paul, Yasmin, Joerg, Sandra, Annabel, Judith, Vero, and Rachel.

En ook heel erg bedankt aan Ron, Marianne, Pim en Tom.

I would like to thank my mum from the bottom of my heart, for her love, support and friendship throughout this considerable DPhil process. I would also like to thank my dad, my brother, my twin and my little sister, my uncle Guy, my aunt Bron and Dean for their tremendous love and support.

And last but not least I would like to thank Frank for being there, for being absolutely incredible and making it all seem worth it in the end.