

Attention regulates the plasticity of multisensory timing

James Heron,¹ Neil W. Roach,² David Whitaker¹ and James V. M. Hanson¹

¹Bradford School of Optometry and Vision Science, University of Bradford, Bradford, UK

²Visual Neuroscience Group, School of Psychology, The University of Nottingham, Nottingham, UK

Keywords: adaptation, asynchrony, audiovisual, temporal perception

Abstract

Evidence suggests that human time perception is likely to reflect an ensemble of recent temporal experience. For example, prolonged exposure to consistent temporal patterns can adaptively realign the perception of event order, both within and between sensory modalities (e.g. Fujisaki *et al.*, 2004 *Nat. Neurosci.*, **7**, 773–778). In addition, the observation that ‘a watched pot never boils’ serves to illustrate the fact that dynamic shifts in our attentional state can also produce marked distortions in our temporal estimates. In the current study we provide evidence for a hitherto unknown link between adaptation, temporal perception and our attentional state. We show that our ability to use recent sensory history as a perceptual baseline for ongoing temporal judgments is subject to striking top-down modulation via shifts in the observer’s selective attention. Specifically, attending to the temporal structure of asynchronous auditory and visual adapting stimuli generates a substantial increase in the temporal recalibration induced by these stimuli. We propose a conceptual framework accounting for our findings whereby attention modulates the perceived salience of temporal patterns. This heightened salience allows the formation of audiovisual perceptual ‘objects’, defined solely by their temporal structure. Repeated exposure to these objects induces high-level pattern adaptation effects, akin to those found in visual and auditory domains (e.g. Leopold & Bondar (2005) *Fitting the Mind to the World: Adaptation and Aftereffects in High-Level Vision*. Oxford University Press, Oxford, 189–211; Schweinberger *et al.* (2008) *Curr. Biol.*, **18**, 684–688).

Introduction

A multitude of studies have shown that our ability to compute estimates of event time depends on an ensemble of preceding sensory and/or motor experience. For example, a recent history of eye movements (Yarrow *et al.*, 2001; Morrone *et al.*, 2005), predictable stimuli (Pariyadath & Eagleman, 2008), repeated exposure to consistent temporal intervals (Behar & Bevan, 1961; Walker *et al.*, 1981) and temporal frequencies (Recanzone, 2003; Johnston *et al.*, 2006; Burr *et al.*, 2007) can have profound effects on our post-adaptation temporal estimates.

A more controversial finding is that perceived temporal order between multisensory signals is also a product of recent experience. Two influential studies independently demonstrated that adaptation to a consistent polarity (e.g. sound leads vision) and magnitude (e.g. 120 ms) of audiovisual asynchrony distorts the post-adaptation perception of temporal order (Fujisaki *et al.*, 2004; Vroomen *et al.*, 2004). Specifically, the physical asynchrony that corresponds to perceptual synchrony (the point of subjective simultaneity; PSS) now resembles the adapting stimuli. In other words, following adaptation, observers require the same polarity of asynchrony in order to perceive the signals as simultaneous (perceptual simultaneity now requires a physical lead of sound over vision).

Subsequent attempts to investigate the magnitude and generality of this effect have produced a variety of conflicting findings. If effect size is expressed as a percentage of the adapting stimulus, reported audiovisual recalibration values are approximately 7% (Vroomen *et al.*, 2004; Keetels & Vroomen, 2007), 13% (Fujisaki *et al.*, 2004), 32% (Harrar & Harris, 2008), 40% (Hanson *et al.*, 2008a) and 50% (Heron *et al.*, 2007). In addition, it has been argued that the effect is both specific (Hanson *et al.*, 2008b) and nonspecific (Keetels & Vroomen, 2007) to the spatial location of the adapting stimuli. The existence of similar effects in the visuotactile and audiotactile domains is more controversial still. Different groups report either similar PSS shifts across all three sensory pairings (Hanson *et al.*, 2008a), a complete absence of recalibration outside the audiovisual domain (Navarra *et al.*, 2007; Harrar & Harris, 2008) or differing levels of visuotactile recalibration (Keetels & Vroomen, 2008; Takahashi *et al.*, 2008).

Such interstudy variability raises the question of whether temporal recalibration effects should be described as mandatory or cognitive in nature. Recently, a mechanism for audiovisual temporal recalibration has been proposed in which the perceived onset times of auditory signals are either speeded up or slowed down (relative to their visual counterparts) following adaptation to auditory lags or leads, respectively (Navarra *et al.*, 2009). This finding suggests that recalibration between sound and vision may have an early peripheral locus, similar to that proposed for audiovisual speech perception (van Wassenhove *et al.*, 2005) and the ‘sound-induced illusory double flash’ phenomenon (Shams *et al.*, 2001).

Correspondence: Dr J. Heron, as above.
E-mail: j.heron2@bradford.ac.uk

Received 23 September 2009, revised 18 November 2009, accepted 19 November 2009

However, in the visual domain, adaptation aftereffects that were once thought to be entirely stimulus-driven have since proved to be remarkably susceptible to the attentional state of the observer. For example, the magnitude of the motion aftereffect (Chaudhuri, 1990; Culham *et al.*, 2000; Rezec *et al.*, 2004; Verstraten & Ashida, 2005) and tilt aftereffect (Spivey & Spirm, 2000; Suzuki, 2001; Montaser-Kouhsari & Rajimehr, 2004; Kanai *et al.*, 2006; Bahrami *et al.*, 2008) show a strong dependence on the focus of observer's attention during adaptation. For the motion aftereffect, the modulatory influence of attention appears to parallel dynamic modulation of related neural activity (Huk *et al.*, 2001; Seiffert *et al.*, 2003).

In the current study, we investigated the interaction between adaptation, selective attention and temporal perception. We show that, during adaptation, the focus of attention had a decisive effect on the magnitude of the aftereffect. Specifically, attending to the relative onset times of auditory and visual stimuli dramatically amplified the effect of repeated exposure to audiovisual asynchrony. Our data show that top-down processing plays a hitherto unidentified role in modulating the perceptual changes induced by recent temporal experience.

Materials and methods

Subjects

Seven trained observers (three authors, four naïve) participated in the first experiment whereas four trained observers (three authors, one naïve) participated in subsequent experiments. All experiments were undertaken with the understanding and written consent of each subject, and were conducted in accordance with The Declaration of Helsinki.

Stimuli

The visual stimulus was a Gaussian blob ($\sigma = 2^\circ$, luminance 100 cd/m^2) presented for one frame (10 ms) at the centre of a Mitsubishi Diamond Pro 2070 22-inch CRT monitor (100 Hz refresh rate, background luminance 50 cd/m^2). The exact moment of presentation was controlled by a ViSaGe Visual Stimulus Generator (VSG; Cambridge Research Systems, UK), which synchronised presentation to the refresh cycle of the monitor. The centre of the stimulus was aligned with the centre of a fixation cross. The auditory stimulus was a 10-ms square-wave windowed burst of bandpass-filtered (200 Hz–12 kHz) white noise (70 dB SPL) delivered binaurally via Sennheiser HD650 linear headphones. For each auditory presentation, the white noise was convolved with the observers' individually recorded head-related transfer functions representing a spatial offset of 0° (i.e. immediately in front of the observer) or $+10^\circ$ (i.e. 10° left of midline). This arrangement produced an auditory stimulus spatially coincident with the visual stimulus ('attend temporal order' and 'attend fixation' conditions; see below) or a 10° spatial disparity between the two ('attend stimuli' condition). The experiment was controlled by custom-written software in MatLab (Mathworks, USA) on a Dell desktop PC.

Procedures

At the start of each experimental session, observers fixated a central fixation cross ($0.2 \times 0.2^\circ$) on the computer monitor. Each experimental session began with an 'adaptation phase', during which observers were exposed to 100 audiovisual stimulus pairs with an asynchrony of

120 ms. The polarity of the adapting stimulus pairs (e.g. 'vision leads sound') was consistent throughout an experimental run. Each stimulus pair was separated by an interstimulus interval (ISI) which varied randomly (with a uniform probability) between 1080 and 1400 ms. Following the adaptation phase, a pause of 2200–2400 ms alerted the observer that the 'test phase' of the experiment was about to commence. Before each test presentation, observers were presented with four 'top-up' stimulus pairs with the fifth stimulus pair being the test stimulus (Fig. 1A). Top-up stimuli were identical to those presented in the preceding adaptation phase (i.e. they had the same polarity and magnitude of asynchrony) and served to maintain the adaptation levels between presentation of test stimuli. The test stimuli themselves were presented at one of seven possible asynchronies: -120 , -80 , -40 , 0 (simultaneous), 40 , 80 and 120 ms, which were randomly interleaved within a method of constant stimuli. In the present study, positive asynchronies refer to a physical lead of sound over vision (Fig. 2A). Observers made un-speeded, binary forced-

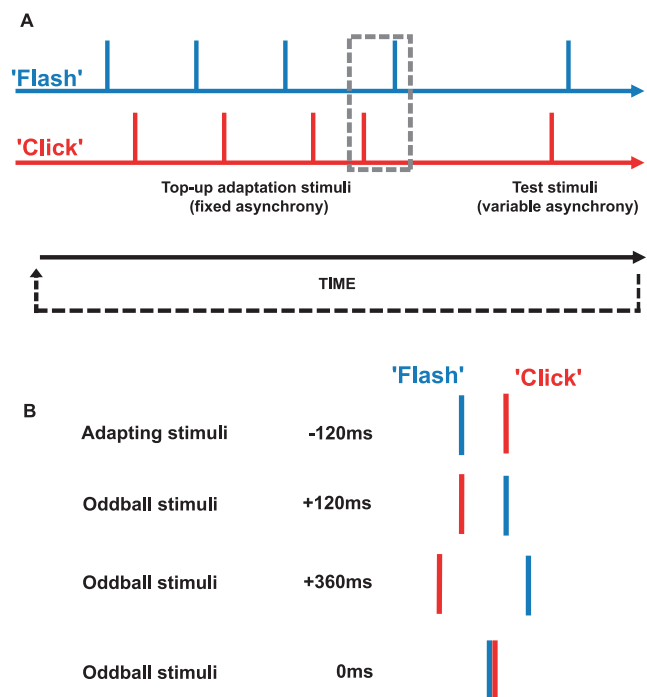


FIG. 1. Schematic showing the relationship between top-up, test and oddball stimuli. (A) An example of the stimuli deployed during top-up and test phases where top-up stimuli comprise a visual lead over sound ('flash' precedes 'click') and test stimuli comprise a small auditory lead over vision. Prior to the top-up phase observers were exposed to 100 stimulus pairs which were physically identical to those shown in the top-up phase. Throughout the adaptation and top-up phases there was a 5% probability that observers would encounter oddball stimuli. In this example the oddball stimulus comprises a reversal of temporal order (for this presentation alone, 'click' precedes 'flash'—see area within grey dashed box). On detection of an oddball, observers made a speeded reaction time response via the keyboard (for details see Procedures). After presentation of the test stimuli observers judged their temporal order by pressing a key which initiated the next top-up and test cycle. (B) A schematic illustrating manipulation of the salience of temporal order oddball stimuli. (i) An example: 120 ms asynchrony between adapting stimuli ('flash' precedes 'click') and (ii) a reversed temporal order oddball stimulus, producing 240 ms asynchrony difference between adaptation and oddball stimuli (see green data in all subsequent figures). (iii) Increasing the asynchrony of the oddball to +360 ms increased its salience by providing a 480 ms difference between adaptor and oddball, whereas (iv) decreasing its asynchrony to 0 ms decreases its salience by providing 120 ms difference between oddball and adaptor (see rightmost and leftmost grey bars respectively in Fig. 4A and B).

choice temporal-order judgments as to 'which modality came first, sound or vision?' and responded via the computer keyboard.

Throughout the course of both adaptation and top-up phases, 'oddball' stimulus pairs were presented with 5% probability. On detection of an oddball stimulus observers were instructed to respond as quickly as possible by pressing a key on a computer keyboard. The time taken for observers to respond to each oddball presentation (i.e. the reaction time; RT) was recorded for off-line analysis. When an oddball stimulus pair was presented during the adaptation phase, that trial was excluded from the total of 100 adapting stimulus pairs. However, when an oddball was presented during the top-up phase the top-up cycle was reset, providing a further four top-up stimulus pairs followed by the test pair. Three different oddball stimulus pairings were used:

1. 'Attend fixation' condition: a sudden reversal in the contrast polarity (black to white) of the central fixation cross.
2. 'Attend stimuli' condition: the visual stimulus was presented at half of its regular size ($\sigma = 1^\circ$ rather than 2°), and the auditory stimulus was convolved with the head-related transfer functions corresponding to a spatial offset of $+10^\circ$.
3. 'Attend temporal order' condition: the polarity of the adapting stimulus pairs was reversed. For example, during adaptation to a 120-ms physical lead of sound over vision, the oddball stimulus pair consisted of a visual stimulus physically leading an auditory stimulus by 120 ms. Thus, observers had to detect a change in asynchrony of 240 ms.

Prior to commencing each experimental sitting, observers were instructed which of the oddball stimuli would be the target oddball and to prepare themselves to respond to this oddball when presented. In this way, observers' attention was directed to a different aspect of the audiovisual stimulus pairs in each of the three conditions. Importantly, all three oddballs were presented with equal probability on each experimental sitting. Thus, all adaptation test and oddball stimuli were physically identical within a given experimental condition. Critically, the only difference between conditions was the focus of observers' selective attention.

In a subsequent control experiment observers detected versions of the temporal order oddballs described above with the difference being that the saliency of the oddball was either halved (120 ms difference between oddball and adapting stimulus: 'attend 120 ms' condition) or doubled (a 480 ms difference: 'attend 480 ms' condition (Figs 1B and 4A and B). This was achieved by keeping the adapting time lag at a constant value (e.g. -120 ms) whilst the asynchrony of the oddball was manipulated to provide either a halving or doubling of the standard 240-ms oddball difference (see Fig. 1B). In the final experiment observers detected an oddball ISI ('attend ISI' condition: Fig. 5A and B) where oddball ISI stimuli comprised of a 400-ms reduction in the ISI separating flash-click adapting stimulus pairs (i.e. 680–1000 ms). These control experiments were otherwise identical to the first experiment.

Each experimental sitting contained 10 presentations at each asynchrony, and each observer completed five experimental sittings in each of the three oddball conditions, making a total of 2100 presentations per observer (10 presentations $\times$ 7 stimulus-onset asynchronies $\times$ 5 experimental runs $\times$ 2 adaptation polarities $\times$ 3 oddball conditions). Whilst the target oddball and adaptation polarity were fixed within each experimental sitting, across sittings the presentation order of each condition was randomised.

For all observers, the percentage of 'sound-first' responses for each condition was plotted as a function of asynchrony and fitted with a logistic function of the form

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

where μ is the audiovisual asynchrony value corresponding to the PSS (the 50% response level on the psychometric function), and θ provides an estimate of temporal order threshold (approximately half the offset between the 27 and 73% response levels). In this way, PSS values were obtained for all observers in all of the conditions (e.g. Fig. 2A). Figures 2B, 4B and 5B plot 'Aftereffect magnitude' as the arithmetic difference between PSS values for each adapting polarity:

$$\text{Aftereffect magnitude} = (\text{PSS}_{\text{adapt A leads V}}) - (\text{PSS}_{\text{adapt V leads A}})$$

This provides a measure of the overall extent of the temporal recalibration observed in each attention condition: values close to zero reflect situations where observers are unaffected by the temporal relationship between the adapting stimulus pairs.

Results

Figure 2A shows a sample of the resultant psychometric functions for each adapting polarity (adapt 'vision first', squares; adapt 'audition first', circles) and each attention condition (attend stimuli, blue; attend fixation, red; attend temporal order, green) for one representative observer. The lateral separation between functions of the same colour serves as an indicator of the effects of adaptation. Taking the example of the 'attend temporal order' condition (in green), it can be seen that adapting to a physical lead (circles) or lag (squares) of audition caused the proportion of sound-first responses to be reduced or increased respectively. As a result, the PSS (the function's midpoint) was shifted in the direction of the adapting stimulus. This reflects the fact that adapting to a physical lead of one modality over another has the effect of necessitating the same temporal relationship for subsequent test stimuli to appear simultaneous. Comparison of the lateral separation between the functions (and associated PSS values) shows clear differences between the different conditions. Specifically, repeated exposure to asynchrony is a far more effective driver of temporal recalibration when selective attention is directed to the temporal structure of the adapting stimuli relative to conditions where observers attend to nontemporal stimulus features ('attend stimuli', blue curves) or independent, non-adapting stimuli ('attend fixation', red curves). It is important to note that these differences arise despite the presentation of physically identical sequences of adapting stimuli in all conditions.

Figure 2B shows 'recalibration magnitude' averaged across observers ($n = 7$); this is computed by taking the arithmetic difference between PSS values for the two adapting polarities for each attention condition (i.e. the separation between functions for the same colour shown in Fig. 2A: see Materials and Methods for details). Thus, if any of the attentional manipulations were able to effectively prevent significant temporal recalibration the height of the bars (Fig. 2B) would be close to zero. Figure 2B shows that this is not the case: although the recalibration magnitude is markedly amplified by attending to temporal order, small effects persist despite diverting attention toward nontemporal stimulus features (blue bar) or independent, nonadapting stimuli (red bar). A repeated-measures ANOVA showed that recalibration magnitude differed significantly between the three conditions ($F_{2,12} = 23.10$, $P < 0.0001$). *Post hoc* analysis corrected for multiple comparisons (Tukey's HSD) showed that the 'attend temporal order' condition significantly differed from both the 'attend fixation' ($P = 0.0002$) and 'attend stimuli' ($P = 0.0002$)

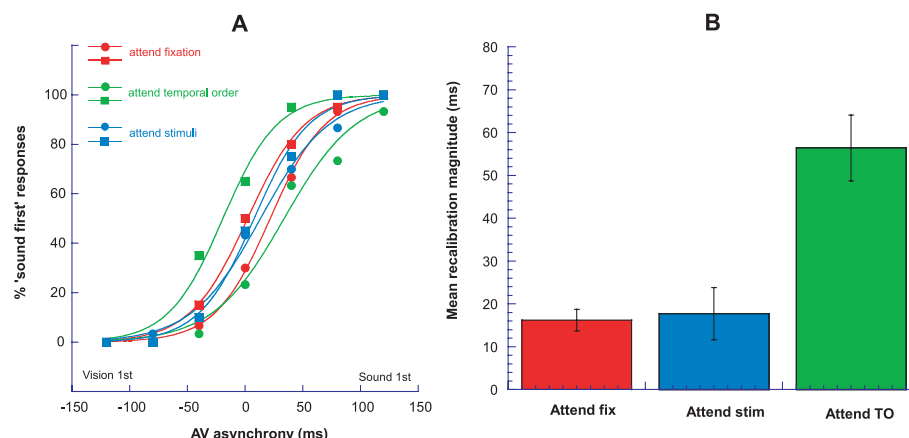


FIG. 2. Attentional modulation of temporal adaptation. (A) Sample psychometric functions from one representative observer showing temporal-order judgments (TOJs) as a function of audiovisual (AV) asynchrony. Square and circle data points represent TOJs following adaptation to 120 ms visual or auditory temporal leads, respectively. Different colours represent conditions in which observers were presented with physically identical stimuli but shifted the focus of their selective attention toward either a fixation cross (red), nontemporal features of the adapting stimuli (blue) or the temporal order of the adapting stimuli (green). For each function, the point of subjective simultaneity (PSS) represents the physical asynchrony value corresponding to perceptual simultaneity (50% 'sound first' responses). (B) Recalibration magnitude averaged across observers ($n = 7$) for each attention condition (fix, fixation cross; stim, stimuli; TO, temporal order). The height of the bars represents the arithmetic difference between PSS values from functions of the same colour shown in A (i.e. opposing adaption asynchrony polarities). The same colour coding applies to both A and B. Error bars represent the SEM between observers.

conditions. The observed recalibration did not differ significantly between the 'attend fixation' and 'attend stimuli' conditions ($P = 0.97$). A single-sample t -test for each condition confirmed that the recalibration magnitude was significantly different from zero (all $df = 6$; 'attend fixation', $P = 0.0006$; 'attend stimuli', $P = 0.028$; 'attend temporal order', $P = 0.0003$). Across conditions, sensitivity to temporal order (i.e. the slope of the psychometric functions) showed no significant differences ($F_{2,12} = 1.51$, $P = 0.26$).

Figure 3A shows median oddball RT from the same representative observer shown in Fig. 2A, whilst Fig. 3B shows median RT averaged across observers. Although RTs in the 'attend temporal order' condition appear slightly slower than those in the 'attend fixation' and 'attend stimuli' conditions (Fig. 3B) these differences failed to reach significance ($F_{2,12} = 1.51$, $P = 0.259$). This finding suggests that task difficulty was well matched between the different conditions. However, some observers reported finding the 'attend temporal order' condition the most cognitively taxing. It could therefore be argued that, for observers to maintain matched performance (in terms of RT; Fig. 3B), different degrees of attentional resources were required

across the different conditions (Brown, 1997). In order to ascertain whether our effects were genuinely attributable to differences in attended stimulus parameters (rather than simple differences in task difficulty) we constructed an additional experiment in which the 'attend temporal order' data from our initial experiment (Figs 2 and 3, green data) were compared with data from two control conditions. In these conditions, the saliency of the temporal order oddball was either halved ['attend 120 ms' condition, in which the asynchrony difference between the oddball and the adapting stimulus was reduced to 120 ms (as against 240 ms in the initial experiment)] or doubled ('attend 480 ms'; see Materials and Methods for details) as shown in Fig. 1B. This manipulation of task difficulty predicts slower and faster reaction times for lower and higher saliency oddballs, respectively. If the extent of the recalibration magnitude tracks any changes in reaction time then differences between conditions in the previous experiment (Fig. 2B) are likely to reflect changes in attentional load between the conditions rather than changes in attentional focus *per se*.

Reaction time and aftereffect magnitude data from this experiment are shown in Fig. 4 (grey bars) alongside their 'attend temporal order'

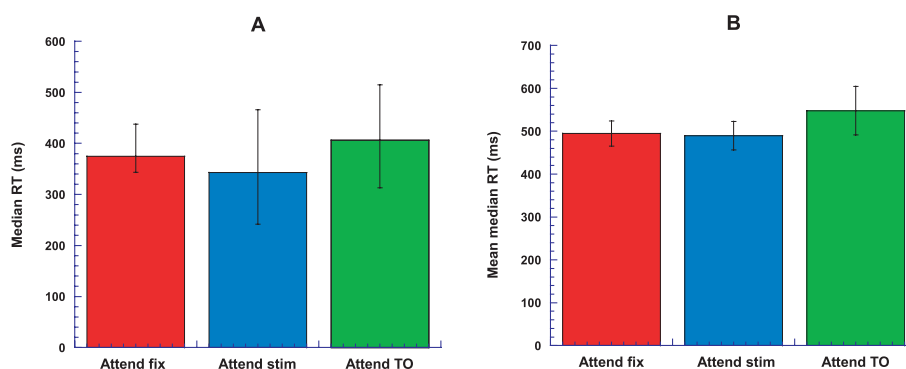


FIG. 3. Task difficulty and selective attention. (A) Median simple RT for the detection of oddball stimuli when observers selectively attended to that oddball. Data shown are from the same representative observer shown in Fig. 2A. The height of the bars represents the perceptual latency for detecting the presence of the oddballs. Error bars represent the interquartile range for each condition. (B) Median RT averaged across observers ($n = 7$). Error bars represent the SEM between observers. For both panels the colour coding is identical to that used in Fig. 2A and B.

counterparts from the initial experiment (green bars; data taken from Figs 2B and 3B). As expected, reaction time was clearly modulated by the salience of the oddball (Fig. 4A); however, the aftereffect magnitude (Fig. 4B) remained constant. Importantly, there was no significant difference between the aftereffect magnitude in either of the control conditions and the ‘attend temporal order’ data from the first experiment (grey and green bars in Fig. 4B; $F_{2,6} = 0.361$, $P = 0.71$). This finding suggests that it was attention to temporal structure, rather than attentional load, that drove the modulation of temporal recalibration in our initial experiment.

A logical question to ask is whether attending to time *per se* is sufficient to drive temporal recalibration or whether relative audiovisual time is the critical perceptual parameter. To test this hypothesis we performed a third experiment where observers attended to the ISI between click–flash pairs. In this setting, selective attention was directed toward a temporal parameter other than that measured during the test phase and only indirectly related to the stimuli themselves (see Materials and Methods for details). If attending to time engages a perceptual mechanism that is insensitive to subdivisions of different temporal features (e.g. perceived temporal frequency, extent, order, onset/offset) then attending to the temporal interval between flash–click pairs should elicit a similar recalibration magnitude to that observed when attending to their temporal order.

Recalibration magnitude data from this experiment are shown in Fig. 5B alongside the accompanying RT data in Fig. 5A where ‘attend ISI’ data (grey bars) are shown alongside ‘attend temporal order’ data from the initial experiment (green bars). Whilst attending to ISI produces significant adaptation effects, it fails to reproduce the recalibration magnitude induced by attending to the temporal order of the flash–click stimulus pairs ($F_{1,3} = 141.29$, $P = 0.00128$). Whilst it remains possible that the small differences in RT between tasks (which just reached significance: $F_{1,3} = 10.86$, $P = 0.04589$) may be a contributing factor in the difference, it appears likely that the process of attending to audiovisual temporal structure (rather than time more generally) is critical in amplifying temporal recalibration.

Discussion

In the current study we have shown that the nervous system’s ability to modulate its representation of time in light of recent experience is highly influenced by the attentional state of the observer. Specifically, when observers selectively attend to the temporal relationship between two stimuli, the effect of repeated exposure to these stimuli is dramatically amplified relative to situations in which selective attention is focused on nontemporal features of the same

stimuli or unrelated, independent stimuli. Control experiments demonstrate that this effect cannot be attributed to differences in attentional load between conditions and that, within the temporal domain, attending to temporal order itself (rather than time more generally) appears to be the most effective driver of temporal recalibration.

These findings offer a plausible explanation for the reasons behind current discord in the literature between the findings of different groups investigating temporal recalibration. The overwhelming majority of these studies strove to engage observers with the adapting stimuli via the detection of oddball stimuli during the adaptation phase. These oddballs typically took the form of an abrupt change in stimulus size (Fujisaki *et al.*, 2004; Takahashi *et al.*, 2008; Navarra *et al.*, 2009), an additional visual stimulus (Vroomen *et al.*, 2004; Keetels & Vroomen, 2007, 2008), auditory pitch (Fujisaki *et al.*, 2004; Navarra *et al.*, 2009) or phonetic detection (Navarra *et al.*, 2005; Vatakis *et al.*, 2008). This approach closely resembles our ‘attend stimuli’ condition and our data suggest that focusing selective attention on nontemporal stimulus features has the unintended consequence of minimising the effect of repeated exposure. Thus, when measuring the extent of temporal recalibration, our findings provide strong evidence for ensuring a close correlation of observer goals (e.g. the focus of selective attention) in both test and adaptation phases. Two studies in which observers were given no attentional instruction during adaptation showed relatively large levels of temporal recalibration (Heron *et al.*, 2007; Hanson *et al.*, 2008a), raising the possibility that, without instructions to the contrary, observers may voluntarily direct selective attention toward the temporal structure of the adapting stimuli (e.g. Fiser & Aslin, 2002; Baker *et al.*, 2004; Turk-Browne *et al.*, 2005) and thus maximise their recalibration levels. It has been proposed that motor actions promote a temporal window of heightened ‘arousal’ (Alexander *et al.*, 2005) and, insofar as attention and arousal are related, this idea may explain the vigorous (often complete) recalibration observed when adapting to asynchrony between our motor actions and their sensory consequences (Liddle & Jackson, 2006; Pesavento & Schlag, 2006; Stetson *et al.*, 2006; Kennedy *et al.*, 2009).

It is important to state that, whilst diverting attention toward the temporal structure of the adapting stimuli had the effect of maximising temporal recalibration, diverting attention away from this parameter did not abolish the basic effect. This finding may be viewed as evidence of a low-level, mandatory component of the effect (one perhaps amenable to amplification by later-stage attentional mechanisms). However, it remains possible that additional manipulations of attention may further amplify the ‘attend temporal order’ effect and

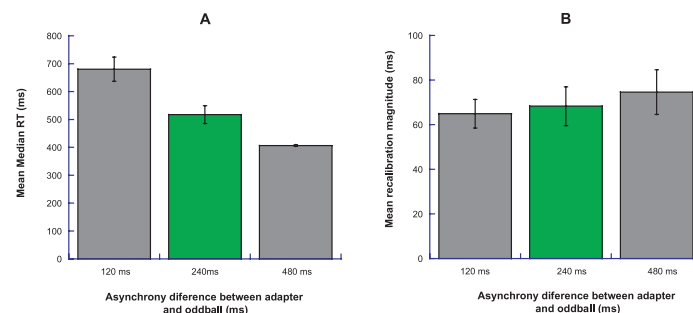


FIG. 4. No change in recalibration despite changes in task difficulty. (A) Grey bars represent observers’ ($n = 4$) mean median RT time for detection of temporal order oddball stimuli under conditions of halved (‘attend 120 ms’) or doubled (‘attend 480 ms’) oddball salience. The green bar represents ‘attend temporal order’ RT data taken from Fig. 3B where the difference between adaptor and oddball was 240 ms. (B) Grey bars represent observers’ mean recalibration magnitude data from the same experiment. The green bar represents the recalibration magnitude data taken from Fig. 2B. For both panels, error bars represent the SEM between observers.

perhaps eradicate effects observed when attention is centered elsewhere. Indeed, several recent studies have argued against low-level interpretations of audiovisual asynchrony perception on the basis that its detection becomes impossible at relatively low stimulus presentation rates, an effect that cannot be ascribed to limitations in the temporal performance of individual visual or auditory systems (Fujisaki & Nishida, 2005, 2007; Benjamins *et al.*, 2008). Moreover, when searching for synchronous targets amongst asynchronous distracters, detection performance declines linearly with the number of distracter stimuli yet performance can be restored via focusing selective attention on an individual audiovisual pair (van de Par & Kohlrausch, 2004; Fujisaki *et al.*, 2006). This pattern of attention-dependent processing has led to the proposal that audiovisual asynchrony perception is mediated by a later-stage or 'mid-level' mechanism that must extract salient auditory and visual features before making temporal cross-correlations across sensory channels (Fujisaki & Nishida, 2005, 2007, 2008), as opposed to earlier detection by specialised low-level sensors. Our findings provide support for this proposal, in that our effects clearly point toward a more complex mechanism that cannot be explained solely on the basis of stimulus-driven factors.

The question arises as to how selective attention interacts with adaptation to mediate temporal recalibration. One parsimonious mechanism would be for attention to simply induce a perceptual expansion of the temporal interval between the auditory and visual stimuli, akin to the attention-dependent expansion of temporal extent (Treisman, 1963; Thomas & Wacaver, 1975; Brown, 1985; Zakay, 1998). This would have the effect of increasing the effective perceptual magnitude of the adapting asynchrony which, within limits, has been shown to promote larger recalibration effects (Fujisaki *et al.*, 2004; Vroomen *et al.*, 2004; Heron *et al.*, 2007). However, attentional expansion of perceived duration scales with the degree of attention allocated to the temporal task, yet our control experiment showed that perceptual load can be manipulated without an accompanying change in recalibration magnitude (Fig. 4A and B). In addition, tuning data from Fujisaki *et al.* (2004) suggest that, in order to produce PSS shifts similar to those from our 'attend temporal order' condition (Fig. 2B), selective attention would need to induce a three-fold expansion in the perceived size of the adapting asynchrony, an unlikely scenario given the much smaller values reported elsewhere (see Brown, 1997 for a review).

An alternative explanation is suggested by evidence from visual experiments in which a consistent temporal relationship promotes

perceptual grouping between otherwise unrelated visual stimuli (Blake & Yang, 1997; Usher & Donnelly, 1998; Lee & Blake, 1999; Jiang *et al.*, 2002; Wallis *et al.*, 2009). This grouping leads to the formation of a wide variety of perceptual 'objects' (Blake & Lee, 2005). Initially, this effect was thought to be dependent on precise temporal synchrony between object elements, but has recently been shown to be maximal when different elements share a common temporal structure (i.e. with repeated exposure, their features are seen to change at times which form a consistent temporal pattern) rather than synchrony *per se* (Guttman *et al.*, 2005, 2007). In the current study, this notion has parallels with the repeated presentation of asynchronous audiovisual pairs which, over time, allow the perceptual system to construct an audiovisual object which is defined by its temporal structure (e.g. Stone, 1998).

We propose that the type of temporal recalibration demonstrated in the current study and elsewhere may in fact reflect an attention-dependent repulsive aftereffect that arises from adaptation to the temporal structure of an audiovisual object. This concept is analogous to effects found in vision and audition where adaptation to a wide variety of high-level stimuli such as geometric shapes (Petersik, 1984; Suzuki, 2001), natural images (Webster & MacLin, 1999; Leopold *et al.*, 2001), voices (Schweinberger *et al.*, 2008) and spatiotemporal patterns (Arnold & Anstis, 1993) cause repulsive aftereffects similar to those observed in the current study. In many cases, such aftereffects are minimised in the absence of attention (e.g. Moradi *et al.*, 2005) and cannot be attributed to adaptation of individual low-level stimulus features (see Leopold & Bondar, 2005; for a review), implying that the effects are a consequence of repeated exposure to the multi-part object itself. For our temporal adaptation effects, selective attention is most likely to intervene by facilitating the extraction of the salient temporal features of our audiovisual object, as demonstrated by studies showing that synchrony/asynchrony judgments are enhanced when observers attend to specific audiovisual pairs amid an asynchronous background (Fujisaki *et al.*, 2006; Fujisaki & Nishida, 2008). The exact processing stage at which attention's influence becomes manifest is open to debate. It could increase signal-to-noise ratios within a later-stage saliency matching mechanism (Fujisaki & Nishida, 2005, 2007, 2008) or alternatively it could boost the salience of the raw sensory signals at much earlier, more peripheral, neural structures (O'Connor *et al.*, 2002; McDonald *et al.*, 2005; Bahrami *et al.*, 2007).

If adapting to the temporal structure of audiovisual pairs does indeed bring about object-based adaptation, it would explain why our effects are amplified by centring attention on temporal order but do not scale

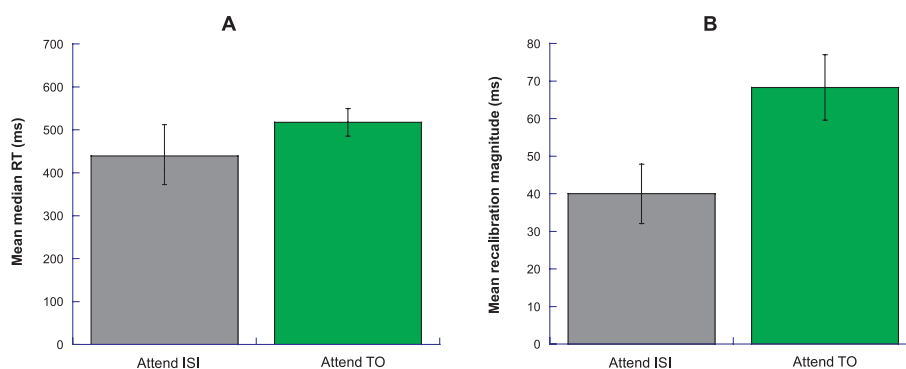


FIG. 5. Attending to ISI vs. attending to temporal order. (A) Grey bar represents mean median RT data for a condition in which observers attended to the ISI separating audiovisual stimulus pairs (as opposed to the within-pair audiovisual asynchrony) by detecting the presence of oddball ISI that were 400 ms shorter than the standard ISI length from the same experiment. The green bar represents 'attend temporal order' RT data taken from Fig. 3B. (B) Grey bar represents mean recalibration magnitude ($n = 4$) from the same experiment. The green bar represents the 'attend temporal order' recalibration data taken from Fig. 2B.

with further increases in the attentional allocation to the same feature (Fig. 4A and B): after reaching a threshold for the perception of a coherent audiovisual object, the object fails to be further established by increases in perceptual load. In addition, if consistent temporal structure is the most effective driver of perceptual grouping, as is the case in the visual domain (Guttman *et al.*, 2005, 2007), then attending to nontemporal stimulus features is less likely to provide a compelling audiovisual object to which observers are able to adapt. This tallies with the much smaller amount of recalibration in this condition (Fig. 2A and B).

The chance probability of repeated temporal correlations between independent events is much lower than the probability of these events occurring in isolation (Guttman *et al.*, 2007). Thus, by monitoring the patterns of change in streams of temporal information (as our observers did in the current study), the nervous system could gain access to a relatively robust form of timing information. Such a strategy could offer a simple solution to the problems faced by the nervous system when combining temporal information from auditory and visual systems beset by differential physical and neural latencies (Spence & Squire, 2003; Heron *et al.*, 2007): while such latencies may cause isolated correlations between independent auditory and visual signals, over time repeated correlations will reliably signal a common external cause.

The nature of the neural mechanisms underpinning distortions of temporal perception remain far from clear. Although human brain imaging studies have made some progress (e.g. Coull *et al.*, 2004; Davis *et al.*, 2009), evidence for a mechanism linking adaptation, temporal processing and selective attention remains sparse. Perhaps the most promising line of enquiry is that represented by a group of recent neurophysiological studies linking temporal perception with the response characteristics of cortical neurons in frontoparietal regions of awake behaving primates (Lebedev *et al.*, 2008; Genovesio *et al.*, 2009; Mita *et al.*, 2009). We believe our findings present testable hypotheses for researchers engaged in these studies. For example, how does the effect of repeated exposure modulate the latency, firing rate or tuning properties of these neurons? Equally, in what fashion are these response properties modulated by top-down influences of the type employed in the current study?

In summary, the results of the current study show that our ability to recalibrate temporal perception in light of recent sensory history is highly influenced by top-down factors. Manipulating observers' attentional state produced large differences in temporal recalibration despite exposure to physically identical stimuli. We propose a mechanism by which selective attention enhances the salience of audiovisual temporal structure, permitting temporal recalibration by high-level pattern adaptation mechanisms. These findings provide new insight into the factors governing the highly flexible process that is human temporal perception.

Acknowledgements

This work is supported by the College of Optometrists, UK, and The Wellcome Trust, UK.

Abbreviations

ISI, interstimulus interval; PSS, point of subjective simultaneity; RT, reaction time.

References

Alexander, I., Thilo, K.V., Cowey, A. & Walsh, V. (2005) Chronostasis without voluntary action. *Exp. Brain Res.*, **161**, 125–132.

- Arnold, K. & Anstis, S. (1993) Properties of the visual channels that underlie adaptation to gradual change of luminance. *Vision Res.*, **33**, 47–54.
- Bahrami, B., Lavie, N. & Rees, G. (2007) Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Curr. Biol.*, **17**, 509–513.
- Bahrami, B., Carmel, D., Walsh, V., Rees, G. & Lavie, N. (2008) Unconscious orientation processing depends on perceptual load. *J. Vis.*, **8**.
- Baker, C.I., Olson, C.R. & Behrmann, M. (2004) Role of attention and perceptual grouping in visual statistical learning. *Psychol. Sci.*, **15**, 460–466.
- Behar, I. & Bevan, W. (1961) The perceived duration of auditory and visual intervals: cross-modal comparison and interaction. *Am. J. Psychol.*, **74**, 17–26.
- Benjamins, J.S., van der Smagt, M.J. & Verstraten, F.A. (2008) Matching auditory and visual signals: is sensory modality just another feature? *Perception*, **37**, 848–858.
- Blake, R. & Lee, S.H. (2005) The role of temporal structure in human vision. *Behav. Cogn. Neurosci. Rev.*, **4**, 21–42.
- Blake, R. & Yang, Y. (1997) Spatial and temporal coherence in perceptual binding. *Proc. Natl. Acad. Sci. U S A*, **94**, 7115–7119.
- Brown, S.W. (1985) Time perception and attention – the effects of prospective versus retrospective paradigms and task demands on perceived duration. *Percept. Psychophys.*, **38**, 115–124.
- Brown, S.W. (1997) Attentional resources in timing: interference effects in concurrent temporal and nontemporal working memory tasks. *Percept. Psychophys.*, **59**, 1118–1140.
- Burr, D., Tozzi, A. & Morrone, M.C. (2007) Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nat. Neurosci.*, **10**, 423.
- Chaudhuri, A. (1990) Modulation of the motion aftereffect by selective attention. *Nature*, **344**, 60–62.
- Coull, J.T., Vidal, F., Nazarian, B. & Macar, F. (2004) Functional anatomy of the attentional modulation of time estimation. *Science*, **303**, 1506–1508.
- Culham, J.C., Verstraten, F.A.J., Ashida, H. & Cavanagh, P. (2000) Independent aftereffects of attention and motion. *Neuron*, **28**, 607–615.
- Davis, B., Christie, J. & Rorden, C. (2009) Temporal order judgments activate temporal parietal junction. *J. Neurosci.*, **29**, 3182–3188.
- Fiser, J. & Aslin, R.N. (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.*, **28**, 458–467.
- Fujisaki, W. & Nishida, S. (2005) Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Exp. Brain Res.*, **166**, 455–464.
- Fujisaki, W. & Nishida, S. (2007) Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Res.*, **47**, 1075–1093.
- Fujisaki, W. & Nishida, S. (2008) Top-down feature-based selection of matching features for audio-visual synchrony discrimination. *Neurosci. Lett.*, **433**, 225–230.
- Fujisaki, W., Shimojo, S., Kashino, M. & Nishida, S. (2004) Recalibration of audiovisual simultaneity. *Nat. Neurosci.*, **7**, 773–778.
- Fujisaki, W., Koene, A., Arnold, D., Johnston, A. & Nishida, S. (2006) Visual search for a target changing in synchrony with an auditory signal. *Proc. Biol. Sci.*, **273**, 865–874.
- Genovesio, A., Tsujimoto, S. & Wise, S.P. (2009) Feature- and order-based timing representations in the frontal cortex. *Neuron*, **63**, 254–266.
- Guttman, S.E., Gilroy, L.A. & Blake, R. (2005) Mixed messengers, unified message: spatial grouping from temporal structure. *Vision Res.*, **45**, 1021–1030.
- Guttman, S.E., Gilroy, L.A. & Blake, R. (2007) Spatial grouping in human vision: temporal structure trumps temporal synchrony. *Vision Res.*, **47**, 219–230.
- Hanson, J.V.M., Heron, J. & Whitaker, D. (2008a) Recalibration of perceived time across sensory modalities. *Exp. Brain Res.*, **185**, 347–352.
- Hanson, J.V.M., Roach, N.W., Heron, J. & McGraw, P.V. (2008b) Spatially specific distortions of perceived simultaneity following adaptation to audiovisual asynchrony. *Perception*, **37**, 27.
- Harrar, V. & Harris, L.R. (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp. Brain Res.*, **186**, 517–524.
- Heron, J., Whitaker, D., McGraw, P.V. & Horoshenkov, K.V. (2007) Adaptation minimizes distance-related audiovisual delays. *J. Vis.*, **7**, 5–1.
- Huk, A.C., Ress, D. & Heeger, D.J. (2001) Neuronal basis of the motion aftereffect reconsidered. *Neuron*, **32**, 161–172.
- Jiang, Y., Chun, M.M. & Marks, L.E. (2002) Visual marking: selective attention to asynchronous temporal groups. *J. Exp. Psychol. Hum. Percept. Perform.*, **28**, 717–730.

- Johnston, A., Arnold, D.H. & Nishida, S. (2006) Spatially localized distortions of event time. *Curr. Biol.*, **16**, 472–479.
- Kanai, R., Tsuchiya, N. & Verstraten, F.A. (2006) The scope and limits of top-down attention in unconscious visual processing. *Curr. Biol.*, **16**, 2332–2336.
- Keetels, M. & Vroomen, J. (2007) No effect of auditory-visual spatial disparity on temporal recalibration. *Exp. Brain Res.*, **182**, 559–565.
- Keetels, M. & Vroomen, J. (2008) Temporal recalibration to tactile-visual asynchronous stimuli. *Neurosci. Lett.*, **430**, 130–134.
- Kennedy, J.S., Buehner, M.J. & Rushton, S.K. (2009) Adaptation to sensory-motor temporal misalignment: instrumental or perceptual learning? *Q. J. Exp. Psychol.*, **62**, 453–469.
- Lebedev, M.A., O'Doherty, J.E. & Nicolelis, M.A. (2008) Decoding of temporal intervals from cortical ensemble activity. *J. Neurophysiol.*, **99**, 166–186.
- Lee, S.H. & Blake, R. (1999) Visual form created solely from temporal structure. *Science*, **284**, 1165–1168.
- Leopold, D.A. & Bondar, I. (2005) Adaptation to complex visual patterns in humans and monkeys. In Clifford, C.W. & Rhodes, G. (eds), *Fitting The Mind to the World: Adaptation and Aftereffects in High-Level Vision*. Oxford University Press, pp. 189–211.
- Leopold, D.A., O'Toole, A.J., Vetter, T. & Blanz, V. (2001) Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.*, **4**, 89–94.
- Liddle, E. & Jackson, S.R. (2006) Recalibrating time: when did I do that? *Curr. Biol.*, **16**, R994–R996.
- McDonald, J.J., Teder-Salejari, W.A., Di Russo, F. & Hillyard, S.A. (2005) Neural basis of auditory-induced shifts in visual time-order perception. *Nat. Neurosci.*, **8**, 1197–1202.
- Mita, A., Mushiaki, H., Shima, K., Matsuzaka, Y. & Tanji, J. (2009) Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nat. Neurosci.*, **12**, 502–507.
- Montaser-Kouhsari, L. & Rajimehr, R. (2004) Attentional modulation of adaptation to illusory lines. *J. Vis.*, **4**, 434–444.
- Moradi, F., Koch, C. & Shimojo, S. (2005) Face adaptation depends on seeing the face. *Neuron*, **45**, 169–175.
- Morrone, M.C., Ross, J. & Burr, D. (2005) Saccadic eye movements cause compression of time as well as space. *Nat. Neurosci.*, **8**, 950–954.
- 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. *Cogn. Brain Res.*, **25**, 499–507.
- Navarra, J., Soto-Faraco, S. & Spence, C. (2007) Adaptation to audiotactile asynchrony. *Neurosci. Lett.*, **413**, 72–76.
- Navarra, J., Hartcher-O'Brien, J., Piazza, E. & Spence, C. (2009) Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proc. Natl. Acad. Sci. U S A*, **106**, 9169–9173.
- O'Connor, D.H., Fukui, M.M., Pinsk, M.A. & Kastner, S. (2002) Attention modulates responses in the human lateral geniculate nucleus. *Nat. Neurosci.*, **5**, 1203–1209.
- van de Par, S. & Kohlrausch, A. (2004) *Visual and Auditory Object Selection Based on Temporal Correlations Between Auditory and Visual Cues*. 18th International Congress on Acoustics, Japan, pp. 2055–2058.
- Pariyadath, V. & Eagleman, D.M. (2008) Brief subjective durations contract with repetition. *J. Vis.*, **8**.
- Pesavento, M.J. & Schlag, J. (2006) Transfer of learned perception of sensorimotor simultaneity. *Exp. Brain Res.*, **174**, 435–442.
- Petersik, J.T. (1984) The perceptual fate of letters in two kinds of apparent movement displays. *Percept. Psychophys.*, **36**, 146–150.
- Recanzone, G.H. (2003) Auditory influences on visual temporal rate perception. *J. Neurophysiol.*, **89**, 1078–1093.
- Rezec, A., Krekelberg, B. & Dobkins, K.R. (2004) Attention enhances adaptability: evidence from motion adaptation experiments. *Vision Res.*, **44**, 3035–3044.
- Schweinberger, S.R., Casper, C., Hauthal, N., Kaufmann, J.M., Kawahara, H., Kloth, N., Robertson, D.M., Simpson, A.P. & Zask, R. (2008) Auditory adaptation in voice perception. *Curr. Biol.*, **18**, 684–688.
- Seiffert, A.E., Somers, D.C., Dale, A.M. & Tootell, R.B.H. (2003) Functional MRI studies of human visual motion perception: texture, luminance, attention and after-effects. *Cereb. Cortex*, **13**, 340–349.
- Shams, L., Kamitani, Y., Thompson, S. & Shimojo, S. (2001) Sound alters visual evoked potentials in humans. *Neuroreport*, **12**, 3849–3852.
- Spence, C. & Squire, S. (2003) Multisensory integration: maintaining the perception of synchrony. *Curr. Biol.*, **13**, R519–R521.
- Spivey, M.J. & Spirn, M.J. (2000) Selective visual attention modulates the direct tilt aftereffect. *Percept. Psychophys.*, **62**, 1525–1533.
- Stetson, C., Xu, C., Montague, P.R. & Eagleman, D.M. (2006) Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, **51**, 651–659.
- Stone, J.V. (1998) Object recognition using spatiotemporal signatures. *Vision Res.*, **38**, 947–951.
- Suzuki, S. (2001) Attention-dependent brief adaptation to contour orientation: a high-level aftereffect for convexity? *Vision Res.*, **41**, 3883–3902.
- Takahashi, K., Saiki, J. & Watanabe, K. (2008) Realignment of temporal simultaneity between vision and touch. *Neuroreport*, **19**, 319–322.
- Thomas, E.A. & Waeaver, W.B. (1975) Cognitive processing and time perception. *Percept. Psychophys.*, **16**, 449–458.
- Treisman, M. (1963) Temporal discrimination and the indifference interval. Implications for a model of the “internal clock”. *Psychol. Monogr.*, **77**, 1–31.
- Turk-Browne, N.B., Junge, J.A. & Scholl, B.J. (2005) The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.*, **134**, 552–564.
- Usher, M. & Donnelly, N. (1998) Visual synchrony affects binding and segmentation in perception. *Nature*, **394**, 179–182.
- Vatakis, A., Navarra, J., Soto-Faraco, S. & Spence, C. (2008) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Exp. Brain Res.*, **185**, 521–529.
- Verstraten, F.A.J. & Ashida, H. (2005) Attention-based motion perception and motion adaptation: what does attention contribute? *Vision Res.*, **45**, 1313–1319.
- Vroomen, J., Keetels, M., de Gelder, B. & Bertelson, P. (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn. Brain Res.*, **22**, 32–35.
- Walker, J.T., Irion, A.L. & Gordon, D.G. (1981) Simple and contingent aftereffects of perceived duration in vision and audition. *Percept. Psychophys.*, **29**, 475–486.
- Wallis, G., Backus, B.T., Langer, M., Huebner, G. & Bühlhoff, H. (2009) Learning illumination- and orientation-invariant representations of objects through temporal association. *J. Vis.*, **9**, 1–8.
- van Wassenhove, V., Grant, K.W. & Poeppel, D. (2005) Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U S A*, **102**, 1181–1186.
- Webster, M.A. & MacLin, O.H. (1999) Figural aftereffects in the perception of faces. *Psychon. Bull. Rev.*, **6**, 647–653.
- Yarrow, K., Haggard, P., Heal, R., Brown, P. & Rothwell, J.C. (2001) Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, **414**, 302–305.
- Zakay, D. (1998) Attention allocation policy influences prospective timing. *Psychon. Bull. Rev.*, **5**, 114–118.