Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 44 (2004) 2875–2884

Vision Research

www.elsevier.com/locate/visres

Sensory uncertainty governs the extent of audio-visual interaction ☆

J. Heron *, D. Whitaker, P.V. McGraw

Department of Optometry, University of Bradford, Bradford BD7 1DP, UK

Received 22 January 2004; received in revised form 23 June 2004

Abstract

Auditory signals have been shown to exert a marked influence on visual perception in a wide range of tasks. However, the mechanisms of these interactions are, at present, poorly understood. Here we present a series of experiments where a temporal cue within the auditory domain can significantly affect the localisation of a moving visual target. To investigate the mechanism of this interaction, we first modulated the spatial positional uncertainty of the visual target by varying its size. When visual positional uncertainty was low (small target size), auditory signals had little or no influence on perceived visual location. However, with increasing visual uncertainty (larger target sizes), auditory signals exerted a significantly greater influence on perceived visual location. We then altered the temporal profile of the auditory signal by modulating the spread of its Gaussian temporal envelope. Introducing this temporal uncertainty to the auditory signal greatly reduced its effect on visual localisation judgements. These findings support the view that the relative uncertainty in individual sensory domains governs the perceptual outcome of multisensory integration.

1. Introduction

The question of how humans integrate sensory information arising from an external multisensory event is currently the focus of intense scientific debate. Examples of interaction between different forms of sensory information, at a number of different neural scales, are ubiquitous in the scientific literature. A classic behavioural study often cited is the McGurk effect (McGurk & MacDonald, 1976) where the perceptual output, derived from visual and auditory information, is the product of a unique combination of the two, in which both vie for perceptual dominance. This effect is often cited as evidence of visual biasing of auditory processing. However, there now exist numerous examples of auditory signals exerting a marked influence on many aspects of visual processing, ranging from phenomenological observations (Shams, Kamitani, & Shimojo, 2000) to the interpretation of visual motion (Meyer & Wuerger, 2001; Sekuler, Sekuler, & Lau, 1997).

Human examples of audio-visual interactions across external space have, thus far, taken the form of visual capture of auditory location or "ventriloquism"—where perceived auditory location is biased toward a temporally synchronous visual cue (Bertelson & Radeau, 1981; Hairston et al., 2003; Howard & Templeton, 1966; Slutsky & Recanzone, 2001). Conversely, if judgements are made in the temporal domain, where auditory thresholds are low relative to their visual counterparts, the perceptual output is dominated by auditory information (Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Recanzone, 2003; Shipley, 1964; Walker & Scott, 1981). Given the relatively poor spatial localisation capability of the auditory system, it is perhaps not surprising that any integrated multisensory percept favours the more accurate system (Welch & Warren, 1980).

Traditional ventriloquism studies have employed conditions where reliable auditory and visual cues signal

---

☆ This work was presented at the 4th International Multisensory Research Forum, Hamilton, Canada, June 2003.

* Corresponding author. Tel.: +44 1274 234652; fax: +44 1274 235570.

E-mail address: j.heron1@bradford.ac.uk (J. Heron).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2004.07.001

different spatial locations. However, the inherent spatial uncertainty associated with auditory processing, relative to visual, means that the visual stimulus usually assumes dominance in multisensory localisation. In order to fully examine the extent of potential auditory influences on perceived visual position, it is necessary to employ a task which generates different levels of relative sensory uncertainty in the visual domain. Recent studies suggest the role of auditory spatial signals in cross-modal spatial localisation depends on the spatial reliability of the visual signal (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003). In the present study we re-visit this issue and ask the question: if an auditory signal is to influence perceived visual position, is it necessary for the auditory signal to have a spatial component? Can the temporal characteristics of an auditory signal influence the perceived location of a moving visual target?

Here we utilised a task evoking a commonly occurring multisensory event: a moving object striking a surface and ''bouncing'' (Fig. 1). In this example, both auditory and visual signals provide information, which may or may not be relevant in localising the spatial coordinates of the bounce position. The auditory cue can be delivered in the temporal domain (i.e. at a specific temporal location relative to the bounce position), whilst the visual judgement is made in the spatial domain (i.e. locating the bounce position). If cues from both modalities are temporally synchronous, the nervous system has been shown to categorise the cues as emanating from the same external event at both neural (Meredith, Nemitz, & Stein, 1987) and behavioural (Bertelson & Radeau, 1981; Sekuler et al., 1997; Slutsky & Recanzone, 2001; Watanabe & Shimojo, 2001) levels. Indeed, this multisensory ''binding'' has been demonstrated in the presence of small levels of temporal disparity between auditory and visual signals (McDonald, Teder-Salejarvi, & Ward, 2001; Meredith et al., 1987) hence the so called ''multisensory temporal integration window''. If the nervous system binds auditory and visual information—despite physical asynchrony between the two—then introducing a salient sound prior to the true bounce position may bias the perceived bounce position. Should this illusory percept be dependent on the multisensory temporal integration window then systematically modulating this asynchrony should influence the ability of auditory signals to bias judgements of perceived position.

Within such a temporal integration window, there still remains the question of how the nervous system allocates perceptual weight to the various modalities contributing to the unified percept. A recent study has proposed a weighting model for visuo-haptic interactions where visual dominance occurs when the variance associated with the visual cue is lower than that associated with the haptic cue (Ernst & Banks, 2002). In the present study we adopt a similar approach, but rather than limiting uncertainty modulation to one modality, we co-vary the relative uncertainty of visual *and* auditory information available to observers. If sensory uncertainty is the critical factor in multisensory integration, then modulating the relative levels of uncertainty between auditory and visual domains should characterise any observed effects.

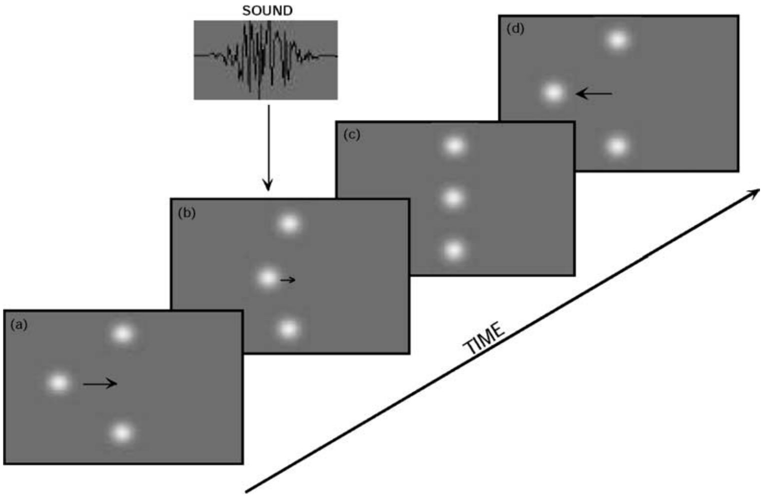

Fig. 1. Panel (a) shows the central Gaussian blob start position. In this particular example the direction of motion is from left to right. As the central Gaussian blob approaches the static references, an auditory stimulus is introduced at a specific temporal location along the motion path of the blob (Panel (b)). The central moving blob reverses its direction of motion, or ''bounces''. Panel (c) depicts the condition where bounce position is in physical alignment with the static reference elements. The temporal asynchrony between bounce position and sound onset can be zero (coincident bounce and sound) or −20, −40, −80 or −160 ms prior to bounce position. After bouncing, the central Gaussian blob then travels back toward its original start position (Panel (d)). The subject's task is to report whether the bounce position occurred to the left or right of the static reference elements.

2. Materials and methods

2.1. Stimuli

The visual stimuli consisted of three identical luminance-defined Gaussian blobs, whose mathematical description is given by

$$L_{\text{mean}} + A * \exp(-(d^2)/2\sigma_{\text{v}}^2) \tag{1}$$

where L_{mean} is the mean luminance of the background, A is the luminance amplitude and σ_{v} is the standard deviation of the Gaussian envelope. σ_{v} could be varied, and ranged from 0.05° to 0.8° in two octave steps. The radial distance from the centre of the Gaussian is denoted by d. The contrast of all stimuli was fixed at 90% Weber contrast.

Visual stimuli were generated using the macro capabilities of the public domain software 'NIH Image 1.61 ppc' (developed at the US National Institutes of Health and available from the Internet by anonymous FTP from zippy.nimh.nih.gov or on floppy disc from the National Technical Information Service, Springfield, Virginia, part number PB95-500195GEI). Stimuli were presented on a 21″. ForMac colour monitor at a mean luminance of 51 $\text{cd}\,\text{m}^{-2}$ and a frame rate of 75 Hz. The non-linear luminance response of the display was linearised using the inverse function of the luminance response as measured with a Minolta CS-100 photometer. The host computer was a Power Macintosh G4. Viewing distance was 43 cm.

The auditory stimuli consisted of Gaussian-windowed broadband white noise bursts. The mathematical description of the temporal window of these stimuli is given by

$$A * \exp(-(t^2)/2\sigma_A^2) \tag{2}$$

where A is the auditory amplitude (sound pressure level (SPL) at a sampling rate of 44.1 kHz), and σ_A is the standard deviation of the Gaussian envelope and is here adopted as a measure of temporal duration (ms). In keeping with the visual stimuli, the value of σ_A could be varied, and was either 2.33 ms—very brief 'click' sound—or 74.56 ms—a longer 'swoosh' sound—a five octave increase in duration. The temporal duration relative to the peak of the Gaussian envelope is denoted by t. Auditory stimuli were generated using customised software, stored as Apple alert sounds and called from within 'NIH Image 1.61 ppc'. This procedure ensured that the auditory stimulus did not interfere with the visual animation sequence. Levels of audio-visual asynchrony were calibrated via simultaneous storage of both signals on a digital oscilloscope. All stimuli were delivered binaurally at equal intensity (70 dB peak SPL) via Sennheiser HD650 linear headphones. Auditory stimuli were measured using a CEL 383 integrating impulse sound level meter.

2.2. Procedures

Visual stimuli were presented in a three-blob Vernier alignment arrangement (see Fig. 1). The centre-to-centre vertical separation of the blobs was a fixed factor of the standard deviation of the Gaussian window (separation was $3.5 \times \sigma_{\text{v}}$). Therefore, separation was proportional to target size, and was fixed within a region where target size is known to limit positional thresholds (Toet, van Eekhout, Simons, & Koenderink, 1987; Whitaker, Bradley, Barrett, & McGraw, 2002). While the subject maintained fixation midway between the two stationary outer reference blobs, the visual target (central blob) was made to translate laterally across the screen, at a fixed velocity of 8.81 deg/s. The start position of the target was randomised, as was its direction of translation (i.e. it could travel either left to right or right to left). Upon reaching one of seven possible locations centred around the point of physical alignment, the central moving target "bounced" and travelled back along its original path, towards its start position (see Fig. 1a–d). An auditory stimulus was introduced either temporally coincident with the "bounce" position of the visual target (audio-visual synchronous condition), or 20, 40, 80 or 160 ms prior to the occurrence of the bounce (audio-visual asynchronous conditions). In the present study the asynchronous conditions all deliver the auditory cue *prior* to the bounce occurrence.

The seven bounce positions and five levels of temporal asynchrony were randomly interleaved within a method of constant stimuli. The subject's task was to report whether the bounce position of the central element occurred before or after reaching the point of perceived alignment with the stationary reference elements. Subjects were instructed to base their judgements solely on the perceived bounce position, and ignore all other cues.

The resulting psychometric functions for each audio-visual temporal asynchrony were fitted with a logistic function of the form

$$y = \frac{100}{1 + e^{-\frac{(x-\mu)}{\theta}}} \tag{3}$$

where μ is the offset corresponding to the 50% level on the psychometric function (point of subjective equality—PSE) and θ provides an estimate of alignment threshold (half the offset between the 27% and 73% response levels on the psychometric function approximately). Each blob size was run separately in conjunction with one of the two auditory stimuli: either a brief "click" or a longer "swoosh" sound, as described above. The responses to the first 10 trials in each run served as a practice period and were not included in the analysis. Data were collected in blocks of 350 presentations (7 bounce positions × 5 auditory temporal locations × 10 repetitions) at each sitting. For each condition, four blocks were added together giving a total of 1400 presentations. This procedure was repeated for each value of σ_V and σ_A. The presentation order of individual blocks was randomised (i.e. for any given block of trials, subjects were equally likely to be presented with any possible combination of σ_V and σ_A).

In an additional experiment we established thresholds for an audio-visual temporal order judgement where subjects estimated the temporal location of an auditory stimulus relative to a visual event—the sudden appearance of a 1.5° diameter circular white disc. We compare thresholds under two conditions: one of small (σ_A = 2.33 ms), and one of large (σ_A = 74.56 ms) auditory temporal spread. Auditory stimuli were generated and presented in a manner identical to that described above. The onset of the auditory stimulus was timed so as to ensure the peak of the Gaussian envelope was coincident with one of seven temporal locations: −150, −100, −50, 0, 50, 100 and 150 ms. Positive values refer to sounds delivered after the abrupt appearance of the disc, with negative values referring to sound before. The seven temporal locations were randomly presented within a method of constant stimuli. The subject's task was to report which came first, the sound (either "click"or "swoosh") or visual stimulus (disc appearance). Subjects responded via the keyboard, which extinguished the visual stimulus and initiated the next trial. The resulting psychometric functions were analysed as described above.

2.3. Subjects

Two of the authors and two naïve subjects participated in the experiments. Subjects had normal visual and auditory function.

3. Results

3.1. Modulating visual uncertainty

Fig. 2a–d represent the results of a unimodal condition in which the perceived bounce position (PSE) was established in the absence of any auditory signal—visual

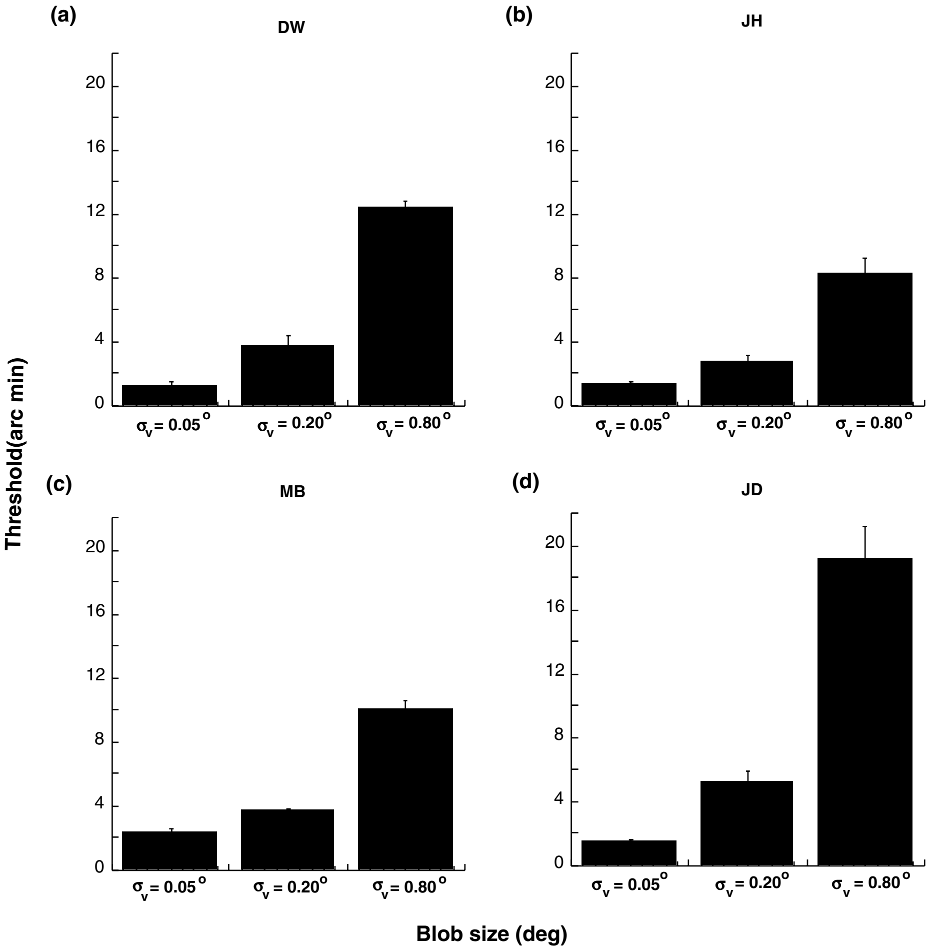

Fig. 2. Thresholds for judgments of bounce position for each of the three-blob sizes σ_V = 0.05°, 0.20° and 0.80°. These reflect the accuracy to which the four observers were able to locate the bounce position. Data represent thresholds obtained in the absence of any auditory stimuli. Error bars indicate the standard deviation of these values.

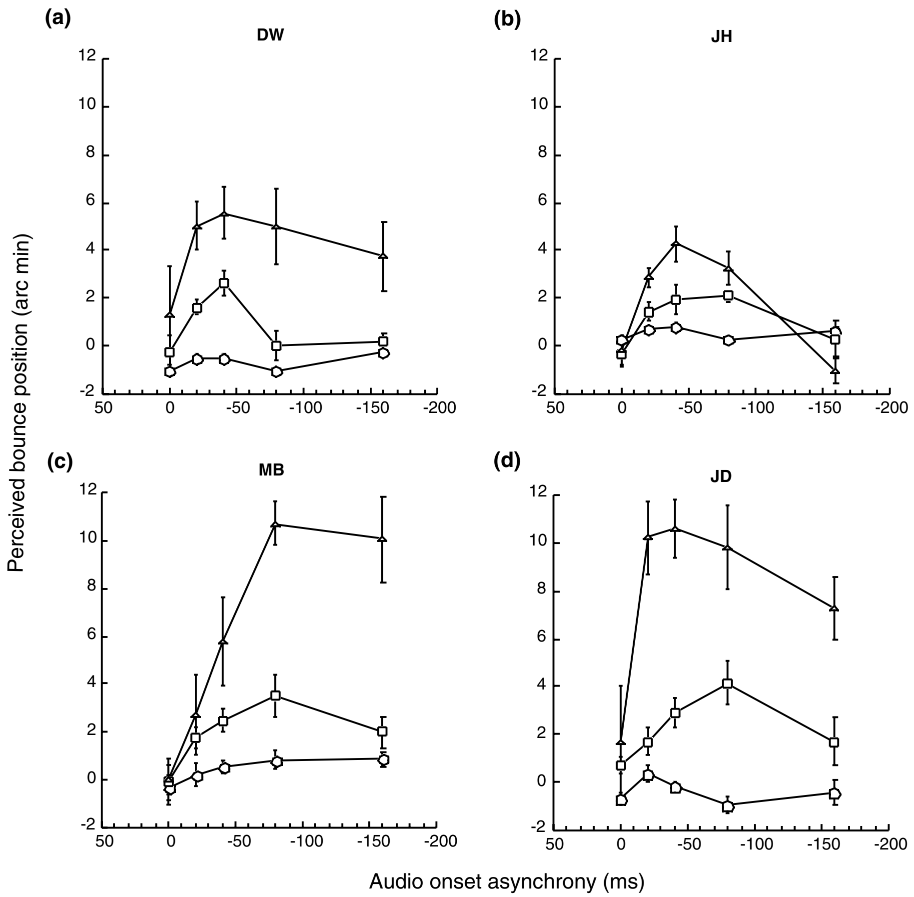

Fig. 3. Perceived spatial location of the bounce position as a function of the audio onset temporal asynchrony for four observers. Ordinate value represents the perceived bounce position relative to the corresponding visual baseline (perceived bounce position in the absence of auditory stimuli). A spatial location of zero represents absence of audio-visual interaction. Positive values on the ordinate represent misperceptions in which the blob appeared to bounce before reaching the static references and vice versa. Errors bars represent one standard deviation either side of the parameter value. Data are shown for three different blob sizes—$\sigma_V = 0.05°$ (circles), 0.20° (squares) or 0.8° (triangles). In all conditions σ_A is held constant at 2.33 ms.

baseline—in order to serve as a comparison with audio-visual conditions. Fig. 2a and b present data for authors JH and DW, whilst 2c and d present data for two naïve observers (MB and JD). The resultant data confirm that positional thresholds do indeed vary systematically with blob size. Specifically, the lowest thresholds are obtained for the smallest blob size ($\sigma_V = 0.05°$) and increase significantly with each two octave increase in blob size. These data are consistent with previous reports (Toet et al., 1987; Whitaker et al., 2002).

Fig. 3 shows cross-modal data for the four observers, where the perceived spatial location of the bounce position is plotted as a function of audio-visual temporal asynchrony. Increasing negative values of audio-visual asynchrony represent an increasing temporal lead of auditory stimuli (sound onset prior to visual bounce). In all plots the temporal characteristic of the auditory stimulus is held constant ($\sigma_A = 2.33$ ms). Data are shown for three different blob sizes—$\sigma_V = 0.05°$ (circles), 0.20° (squares) or 0.80° (triangles). In all panels, the ordinate value represents the perceived bounce position relative to the corresponding visual baseline (perceived bounce position in the absence of auditory stimuli). This highlights the audio-visual interaction effect and facilitates comparison across observers. Thus, an absence of audio-visual interaction results in data points at, or close to, 0 on the ordinate. More positive values on the ordinate represent perceived bounce positions which are closer to that signalled by the auditory stimulus.

For all subjects, the data show that for the audio-visual synchronous condition (an audio onset asynchrony of 0) there is little or no shift in perceived bounce position. For the smallest blob size ($\sigma_V = 0.05°$) perceived bounce position remains relatively constant despite varying levels of audio onset asynchrony (circles in Fig. 3a–d). However, as blob size increases, the auditory signal begins to exert an influence on perceived bounce position (squares and triangles). For example, Fig. 3a–d shows that for all observers maximal auditory influence

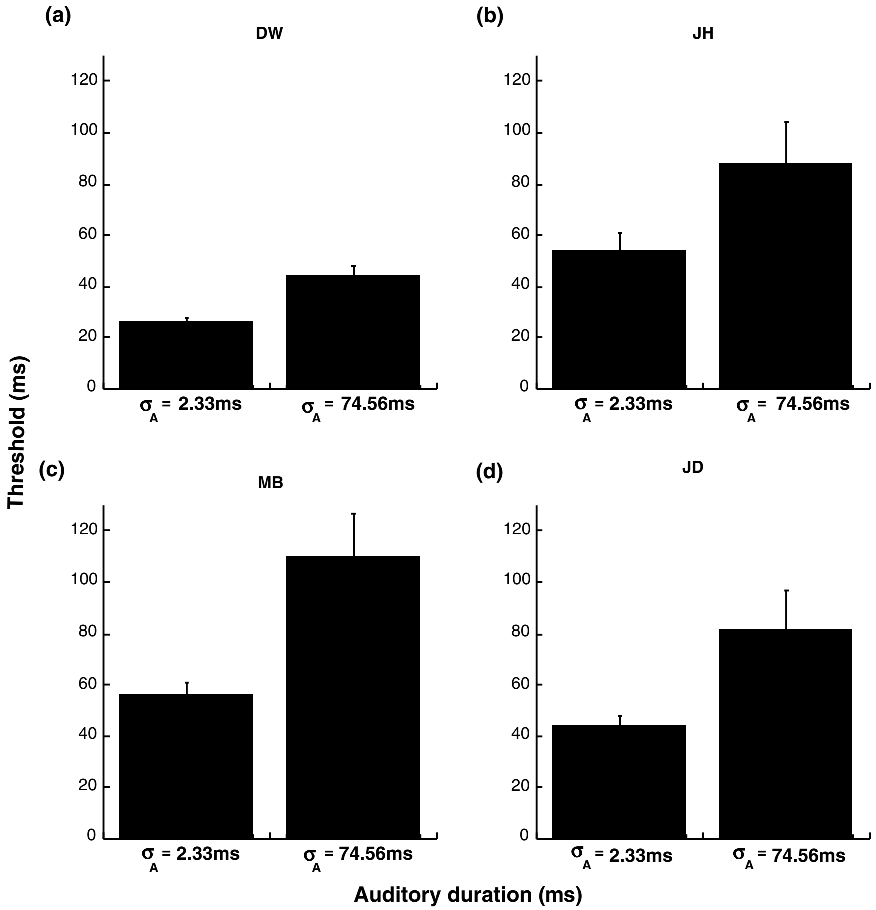

Fig. 4. Thresholds for judging the temporal location of a sound relative to the abrupt occurrence of a visual stimulus by means of a temporal order judgement. Thresholds are shown for four observers under two conditions: one of short (σ_A = 2.33 ms), and one of long (σ_A = 74.56 ms) auditory temporal duration (σ_A is the standard deviation of the Gaussian auditory envelope and is here adopted as a measure of temporal duration).

is evident at asynchronies between −40 and −80 ms. Further increases in audio onset asynchrony beyond this value result in a diminished effect, with estimates of bounce position returning toward their visual baseline values. The relationship between audio onset asynchrony and perceived bounce position allows quantification of the temporal tolerance for audio-visual integration. Within this temporal integration window, auditory signals cause observers to report the bounce position earlier than its physical occurrence. In other words, the perceived bounce position is pulled towards the location occupied by the visual stimulus at the time of the auditory event. This is reflected as a positive shift in perceived bounce position, as shown in Fig. 3a–d. At the intermediate blob size (σ_V = 0.20°, squares), the maximal shift in perceived bounce position is greatest for the naïve observers (MB and JD). As blob size is further increased (σ_V = 0.80°, triangles) the influence of auditory signals on perceived bounce position becomes greater. Whilst the effects are greater for the largest blob size, they either plateau or dissipate with increasing levels of audio onset asynchrony.

3.2. Modulating auditory uncertainty

Fig. 4a–c shows results from the audio-visual temporal order judgement task—where observers were asked to report the temporal location of an auditory relative to a visual event. Although the four observers show differing absolute level of performance, all show that the ability to temporally localise the sound relative to a visual event is dependent on the temporal duration of the auditory stimulus. Increasing σ_A from 2.33 to 74.56 ms induces an approximately twofold increase in threshold.

We now ask the reverse question to the one posed earlier: Does modulating the temporal uncertainty of the auditory stimulus influence the extent of audio-visual interaction? We adopt a similar approach to that of the first experiment, but here we increase the *temporal* duration of the auditory envelope as a means of manipulating its *temporal* uncertainty.

In the first experiment, the maximal auditory influence was found for the largest blob size (σ_V = 0.80°) in conjunction with a brief (σ_A = 2.33 ms) auditory signal.

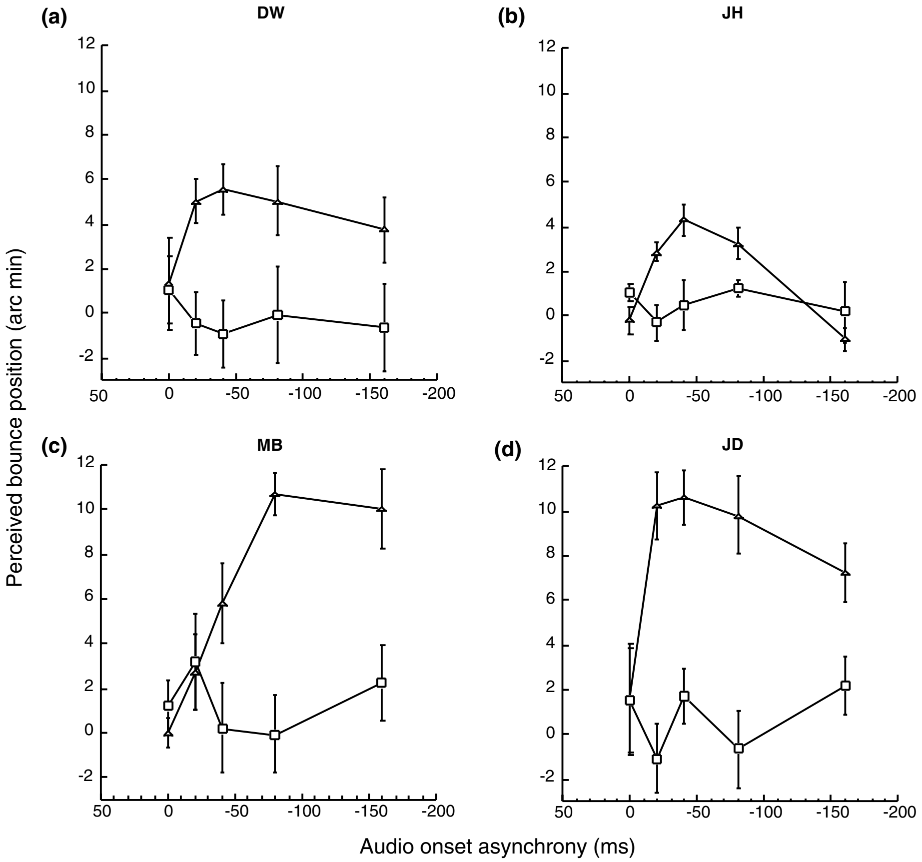

Fig. 5. Perceived spatial location of the bounce position as a function of the audio-visual temporal asynchrony for four observers. Ordinate value represents the perceived bounce position relative to the corresponding visual baseline (perceived bounce position in the absence of auditory stimuli). Positive values on the ordinate represent misperceptions in which the blob appeared to bounce before reaching the static references and vice versa. Errors bars represent one standard deviation either side of the parameter value. Data are shown for two conditions: $\sigma_A = 2.33$ ms (triangles—data taken from Fig. 3) and $\sigma_A = 74.56$ ms (squares). Blob size ($\sigma_v = 0.80°$) is held constant.

We now examine how this effect changes when the duration of the auditory temporal envelope is increased to $\sigma_A = 74.56$ ms, whilst maintaining a fixed level of visual uncertainty ($\sigma_v = 0.80°$). Fig. 5 compares perceived bounce position for short (triangles, taken from Fig. 3) and long (squares) auditory duration. Data for the longer duration (squares) show a distinct lack of audio-visual interaction when compared with its more transient counterpart (triangles).

4. Discussion

The results of the first experiment show that when blob size is small, and therefore visual positional accuracy is high, the introduction of auditory signals have little or no influence on visual judgements of position, irrespective of audio-visual asynchrony. This indicates that when one modality, in this instance vision, operates under conditions of high certainty, the perceptual output is relatively impervious to information from alternative sources. However, as visual positional uncertainty increases, via increases in blob size, the perceived bounce position becomes vulnerable to auditory influence. Specifically, the central blob is perceived to have bounced before it reaches its physical bounce position, presumably pulled by auditory signals delivered prior to the actual blob bounce. Clearly, as visual uncertainty increases, the extent of auditory influence grows correspondingly. Our final experiment revealed an analogous effect, in that for a fixed level of visual uncertainty, modulating auditory uncertainty also determines the extent of multisensory interaction.

These effects are in qualitative agreement with the findings of recent studies examining the integration of information derived from vision and touch (Ernst & Banks, 2002) and vision and audition (Alais & Burr, 2004; Battaglia et al., 2003). These studies manipulated the reliability of the visual signal, with the result that observers based their spatial judgements more upon the signal defined by touch (Ernst & Banks, 2002) or audition (Alais & Burr, 2004; Battaglia et al., 2003). In all three studies, human psychophysical data were compared to the performance of an "ideal" human observer seeking to arrive at a maximum-likelihood estimate derived from multiple estimates of an external physical property. This is thought to be achieved by a linear combination of multiple estimates (either within or between modalities) whose perceptual weight is proportional to the inverse of their variance (Clark & Yuille, 1990). Whilst this model closely predicted multisensory interactions for both Ernst and Banks (2002) and Alais and Burr (2004), Battaglia et al. (2003) found a residual dominance of vision over audition for spatial judgements.

The present study, together with those discussed above, highlights the key role of sensory uncertainty in determining the perceptual weight allocated to respective cues during multisensory integration. With regard to audio-visual positional interactions, our findings demonstrate an auditory influence on perceived visual position when the auditory cue contains only temporal (as opposed to spatial) information. Furthermore, covarying uncertainty in both visual and auditory domains demonstrates how the nervous system can either abolish (Fig. 3) or restore (Fig. 5) visual dominance depending on the relative reliability of sensory information available.

Ernst and Banks (2002) discuss the possibility that relative levels of sensory uncertainty may become irrelevant in circumstances where the individual signals provide such discrepant information to make combination impossible (Ernst & Banks, 2002). Our data show just such an effect. Fig. 3 indicates that the influence of the auditory signal upon the visual judgement dissipates once the audio-visual asynchrony becomes large, with the result that the perceived bounce position returns to baseline levels. In essence, observers are able to ignore any cue which is sufficiently discrepant to make it irrelevant to the judgement. This effect is also apparent in Fig. 3 (squares and triangles), with the suggestion (for three out of four observers) that the temporal window of asynchrony within which an interaction occurs is itself dependent upon the relative uncertainty of the visual and auditory signals.

In the present study we chose *physical* audio-visual synchrony as our zero point on the audio onset asynchrony scale. However it could be argued that *perceptual* synchrony might be more appropriate given the propsed ~50 ms differential neural latency between sound and vision suggested in neurophysiological studies (Meredith et al., 1987). However, recent audio-visual temporal order studies have found that perceptual synchrony, or point of subjective simultaneity (PSS), is elicited when auditory stimuli have only a very small delay relative to visual stimuli. When the distance from the observer is negligible, the magnitude if this delay has been placed anywhere from 5 to 13 ms (Lewald & Guski, 2004; Spence, Baddeley, Zampini, James, & Shore, 2003; Sugita & Suzuki, 2003), typically having a relatively large degree of variance associated with it. Indeed, evidence presented by Stone and colleagues (Stone et al., 2001) suggests each observer may posses their own individual PSS value varying from +20 (vision first) to −150 ms (audition first). Indeed, our own PSS estimates—extracted from the psychometric functions corresponding to the $\sigma_A = 2.33$ ms threshold estimates in Fig. 4—show no systematic deviation from zero (DW: −1 ms, JH: +10 ms, MB: +7 ms and JD: −3 ms).

Several factors may have limited the overall magnitude of our effects. Firstly, the auditory stimuli used in our experiments were internalised along the interaural axis and thus contained no spatial information. The nervous system may allocate greater perceptual weight to auditory signals whose external spatial location coincides with that of the bounce location (as the resultant spatio-temporal overlap may carry greater ecological validity). Secondly, our observers were instructed to base their judgements solely on the perceived bounce position. Our aim was to investigate *mandatory* influences on perceived visual position. It is conceivable that had observers attempted to attend to both modalities, auditory signals may have assumed greater perceptual weight. Such a consideration may help to explain the difference in effect size between naïve (MB and JD in Fig. 3) and trained (DW and JH in Fig. 3) observers. This difference does not appear to be a consequence of significantly degraded visual localisation thresholds in the naïve observers (Fig. 2). A more likely explanation is that our naïve observers display a greater tendency to bind the auditory and visual events into a unified percept since they were unaware of the possible temporal disparity which formed the basis of the present study.

A modular approach to perception has dominated the scientific literature for nearly two decades (Fodor, 1983). The concept of a distinct set of units or modules, each fully devoted to a specific function with their own separate, dedicated neural hardware has provided an invaluable framework for investigating the visual system (Fodor, 1983; Moutoussis & Zeki, 1997; Nakayama, 2001; Zeki et al., 1991). However, the limitations of this view of visual processing must be considered, since it is clear that an amalgamation of the outputs of such modules must take place in order to form the rich (and usually singular) perception of the external world which we experience (Burr, 1999; Lennie, 1998). Within the domain of vision, many examples of interaction between ostensibly independent sources of information have been highlighted. This applies to the judgement of depth (Jacobs, 2002; Landy, Maloney, Johnston, & Young, 1995; Young, Landy, & Maloney, 1993) position (McGraw, Whitaker, Badcock, & Skillen, 2003; Landy & Kojima, 2001; Rivest & Cavanagh, 1996; Whitaker et al., 2002) and orientation (Dakin, Williams, & Hess, 1999; Morgan & Baldassi, 1997; Skillen, Whitaker, Popple, & McGraw, 2002). Recent studies have extended evidence for this type of integrative behaviour into the multisensory arena (King & Calvert, 2001; Shimojo & Shams, 2001) and our present findings shed light on the factors which govern the extent of such effects.

The results are supportive of a perceptual framework where the degree of relative uncertainty in different sensory domains dictates whether the overall perceptual output is derived from one modality alone, or results from multisensory integration. Such an approach would clearly offer ecological benefits. Since real events are often associated with multiple sources of sensory information, each of which tend to be in spatio-temporal register, combination of these sources might serve to enhance the perceptual response to the event (Frassinetti, Bolognini, & Ladavas, 2002; Lovelace, Stein, & Wallace, 2003; Stein, Meredith, Huneycutt, & McDade, 1989). However, in conditions where the various sources of information each provide conflicting estimates of a given stimulus feature, it is essential that any resultant motor commands are based on the most reliable sensory cue available (Jacobs, 2002).

It has been stated that multisensory interactions obey a "Modality Appropriateness" hypothesis (Welch & Warren, 1980), whereby visual dominance in spatial tasks is attributed to the superior localisation accuracy of the visual system relative to other modalities. Under such conditions the visual system could be viewed as a distinct processing module, with auditory signals exerting little or no influence on perceived visual position. However, recent studies suggest a greater level of flexibility in multisensory interaction (Battaglia et al., 2003; Ernst & Banks, 2002; Ernst, Banks, & Bulthoff, 2000; van Beers, Wolpert, & Haggard, 2002). The present study confirms the need for caution when applying the "Modality Appropriateness" to more generalised multisensory interactions: when the nervous system deems visual information to be relatively reliable, it maintains the characteristics of a modular system in which visual information assumes perceptual dominance. On the other hand, when visual uncertainty is high, the output is based on information from alternative sources and auditory signals begin to assert their influence. Degrading auditory reliability restores the dominance of vision. Thus, unless a sensory signal can be discounted as being unconnected with an external event, multisensory perceptual output should be considered as being fundamentally integrative in nature. The mechanism proposed by the "Modality Appropriateness" hypothesis would appear to represent just one extreme on an integrative continuum, where the extent of any interaction is governed by the relative reliability (or uncertainty) of the sensory signals contributing to perception.

Acknowledgments

Author P.V. McGraw is supported by The Wellcome Trust. We would like to thank Dr Kirill Horoshenkov for technical assistance. This work was supported in part by The Visual Research Trust.

References

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

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.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory–visual spatial discordance. *Perception and Psychophysics, 29*, 578–584.

Burr, D. (1999). Vision: modular analysis—or not? *Current Biology, 9*, 90–92.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer.

Dakin, S. C., Williams, C. B., & Hess, R. F. (1999). The interaction of first- and second-order cues to orientation. *Vision Research, 39*, 2867–2884.

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., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3*, 69–73.

Fodor, J. (1983). *Modularity of mind: an essay of faculty psychology*. Cambridge, MA: MIT Press.

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

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience, 15*, 20–29.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: Wiley.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6*, 345–350.

King, A. J., & Calvert, G. A. (2001). Multisensory integration: perceptual grouping by eye and ear. *Current Biology, 11*, 322–325.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A, 18*, 2307–2320.

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.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27*, 889–935.

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.

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*, 447–453.

McDonald, J. J., Teder-Salejarvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science, 292*, 1791.

McGraw, P. V., Whitaker, D., Badcock, D. R., & Skillen, J. (2003). Neither here nor there: localizing conflicting visual attributes. *Journal of Vision, 3*, 265–273, http://journalofvision.org/3/4/2/ doi:10.1167/3.4.2.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

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

Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport, 12*, 2557–2560.

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

Morgan, M. J., & Baldassi, S. (1997). How the human visual system encodes the orientation of a texture, and why it makes mistakes. *Current Biology, 7*, 999–1002.

Moutoussis, K., & Zeki, S. (1997). Functional segregation and temporal hierarchy of the visual perceptive systems. *Royal Society London Series B, 264*, 1407–1414.

Nakayama, K. (2001). Modularity in perception, its relation to cognition and knowledge. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 737–759). New York: Blackwell Publishers Ltd.

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

Rivest, J., & Cavanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research, 36*, 53–66.

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

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*, 788.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology, 11*, 505–509.

Shipley, T. (1964). Auditory flutter driving of visual flicker. *Science, 145*, 1328–1330.

Skillen, J., Whitaker, D., Popple, A. V., & McGraw, P. V. (2002). The importance of spatial scale in determining illusions of orientation. *Vision Research, 42*, 2447–2455.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport, 12*, 7–10.

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

Stein, B. E., Meredith, A. M., Huneycutt, S., & McDade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *Journal Of Cognitive Neuroscience, 1*, 12–24.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society London Series B, 268*, 31–38.

Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: implicit estimation of sound-arrival time. *Nature, 421*, 911.

Toet, A., van Eekhout, M. P., Simons, H. L., & Koenderink, J. J. (1987). Scale invariant features of differential spatial displacement discrimination. *Vision Research, 27*, 441–451.

van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology, 12*, 834–837.

Walker, J. T., & Scott, K. J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1327–1339.

Watanabe, K., & Shimojo, S. (2001). When sound affects vision: effects of auditory grouping on visual motion perception. *Psychological Science, 12*, 109–116.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.

Whitaker, D., Bradley, A., Barrett, B. T., & McGraw, P. V. (2002). Isolation of stimulus characteristics contributing to Weber's law for position. *Vision Research, 42*, 1137–1148.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research, 33*, 2685–2696.

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *Journal of Neuroscience, 11*, 641–649.