Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Brain Research 19 (2004) 185–194

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/cogbrainres

Research report

No direction-specific bimodal facilitation for audiovisual motion detection

David Alais*, David Burr

Istituto di Neurofisiologia del CNR, Via G. Moruzzi, 1, Pisa 56125, Italy

Accepted 18 November 2003

Abstract

After several decades of unimodal perceptual research, interest is turning increasingly to cross-modal interactions. At a physiological level, the existence of bimodal cells is well documented and it is known that correlated audiovisual input enhances localisation and orienting behaviours. Audiovisual perceptual interactions have also been demonstrated (e.g., the well-known McGurk effect). The present study explores motion perception and asks whether correlated audiovisual motion signals would be better detected than unimodal motions or bimodal motions in opposing directions. Using a dynamic random-dot field with variable motion coherence as a visual stimulus, together with an auditory motion defined by a stereo noise source smoothly translating along a horizontal trajectory, we find that correlated bimodal motion yields only a slight improvement (approximately a square root of two advantage) in detection threshold relative to unimodal detection. The size of this benefit is consistent with a statistical advantage rather than a bimodal facilitation account. Moreover, anticorrelated bimodal motion showed the same modest improvement, again speaking against linear summation but consistent with statistical combination of visual and auditory signals. These findings were replicated in peripheral as well as in central vision, and with translating visual objects as well as with spatially distributed visual motion. The superadditivity observed neurally (especially in deep-layer superior collicular cells), when weak unimodal signals are combined in bimodal cells does not apply to the detection of linear translational motion.

Theme: Sensory systems
Topic: Visual (and auditory) psychophysics and behavior

Keywords: Visual psychophysics; Auditory psychophysics; Motion detection; Crossmodal interaction; Bimodal; Audiovisual

1. Introduction

Many objects in the external environment are represented in two or more sensory modalities. Touch and vision are commonly co-activated, when objects are taken in hand and inspected. Audition and vision are also frequently activated by the same stimulus event (the sight and sound of speeding cars is a common example). Despite the modularity of our sensory systems, we perceive a unified and coherent world. Indeed, by synthesising complementary information, we enhance the likelihood that our internal perceptions will accurately reflect external realities [7], enabling us to respond more rapidly and appropriately. Two of the interesting questions which arise from this synthesis are: how is information combined across modalities; and does the perceptual system capitalise on complementary information about the same stimulus to improve its performance. The experiments we present deal primarily with the latter question. Specifically, we ask whether the ability to detect movement is improved when that movement is represented in both auditory and visual modalities.

The combination of information across senses has been heavily researched at the neurophysiological level [30], with particular focus on the superior colliculus. Its deep layers contain many ‘multisensory’ neurons–neurons that receive unimodal sensory input from more than one source. Multisensory cells may be bimodal, or even trimodal, with audiovisual bimodal cells a common variety. These are arranged in a topographical representation of external space and have separate but overlapping auditory and visual receptive fields so that they respond to audiovisual input from a single location [35]. Although they can be driven unimodally, they exhibit a strong non-linear response known as “superadditivity” [12,15] when driven bimodally by spatiotemporally correlated audiovisual input.

* Corresponding author. Current address: Auditory Neuroscience Laboratory, Department of Physiology (F13), University of Sydney, NSW 2006, Sydney, Australia. Tel.: +61-2-9351-7615; fax: +61-2-9351-2058.
E-mail address: alaisd@physiol.usyd.edu.au (D. Alais).

0926-6410/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2003.11.011

Behavioural and attentional studies have demonstrated that cross-modal interactions do indeed occur. Behaviourally, it known that bimodal superadditivity improves orienting behaviours such as eye movements [8,30,32,37] and aids stimulus localisation [1,31]. This likely reflects response enhancement to correlated bimodal stimuli creating a salient peak on collicular topography. Several studies have also confirmed the physiological observation that response enhancement (superadditivity) is maximal when auditory and visual inputs arrive synchronously [12,16], although for perceptual tasks the temporal window within which auditory and visual stimuli can be phenomenally integrated is rather broad [13,21]. In contrast, discordant stimuli lead to ‘‘response depression’’ [11,15] and a corresponding decrease in efficiency of orienting behaviours [32,37]. It has also been shown that attending to a particular spatial location for visual stimuli improves task performance for auditory stimuli (and vice versa) in that location [27,28], suggesting a linked audiovisual topography. This is consistent with the fact that the superior colliculus (containing multi-modal cells) is strongly implicated in orienting to salient stimuli, whether overtly with eye movements or covertly with attention [6,26].

Apart from its role in orienting, the superior colliculus also has strong reciprocal links, via the pulvinar, with middle-temporal (MT) cortical area [29]. MT is an area specialised for processing visual movement and activity in this area is strongly correlated with visual motion perception [2,3]. Outputs from MT project directly to area VIP where they combine with input from auditory areas to create bimodal cells with strong motion selectivity [5,9,14]. Based on the evidence for strong audiovisual interactions in sensory processing, both at an early, subcortical level as well as at higher, motion-specialised cortical areas, we conducted experiments to examine whether sensitivity to bimodally represented movement might be improved relative to unimodal baselines when that movement is spatiotemporally concordant in both audition and vision. In particular, we focused on thresholds for motion detection, as the neurophysiological evidence suggests that response enhancement should be stronger when the unimodal stimuli are weak [15]. On this principle, unimodal stimuli too weak to be detected alone could conceivably become detectable when part of a correlated bimodal stimulus. We therefore measured motion detection thresholds unimodally for vision and for audition, and then again when the stimuli were presented together as a bimodal motion stimulus. We compared conditions in which the auditory and visual components were either matched in direction (correlated) or were opposed (anticorrelated). This was repeated for visual motion in central and in peripheral vision, and for visual stimuli that were spatially distributed or were a spatially localised object. The results show no evidence of a facilitative audiovisual interaction for detection of linear translations, whether in the central or peripheral field.

2. Materials and methods

2.1. Stimuli

The auditory stimuli were created digitally at a sampling rate of 65 kHz and played over loudspeakers (Yamaha MSP5) which lay in the same plane as the video monitor, 45 cm from the observer, and ± 30 cm from the monitor’s centre. The sound was produced by low-pass filtering white noise using a 5th-order Butterworth filter with a cut-off frequency of 2 kHz. Auditory movement was created by playing the filtered signal in stereo and varying the magnitude and sign of interaural time differences so that a diffuse point source was heard to move across the observer’s midline from -20° to $+20^\circ$ (or vice versa) in azimuth over a period of 0.67 s. This resulted in a compelling sense of auditory movement at a constant sound level of 72 dB(A) at the listening position. The strength of the auditory movement signal was manipulated by diluting it with masking noise, although the relative intensity of the two sounds were controlled to keep the total intensity constant. The masking noise had the same spectral characteristics as the motion component, being an independent white noise signal that underwent the same filtering process. The noise component was played in stereo with no interaural time difference and matched binaural phase. The relative amplitudes of the signal and noise sources were manipulated to vary motion strength (or ‘coherence’), which could vary from a value of 0 (no motion signal, only static auditory noise) to a value of 1 (only motion signal). Auditory motion strength is thus the proportion of total audio amplitude represented by the motion component.

The visual stimuli in the first experiments were random dot kinematograms [25] comprised of 100 dark and 100 light dots (0.8° visual angle) arrayed within the full screen of the computer monitor (subtending $50^\circ*38^\circ$) and were redrawn at a rate of 60 Hz. In a similar vein to audio motion strength, the strength of visual movement was manipulated by varying the proportion of dots which carried a motion signal. This subset of dots was displaced uniformly in the signal direction (to the left or right), with a new subset randomly chosen after each frame to carry the subsequent displacement. Drawing a new sample of motion dots each frame prevents subjects tracking individual dots. The other, ‘noise’ dots were allocated new random locations each frame so that they jumped about incoherently and contained no global motion signal. There was no fixation point. As with the auditory signal, the visual stimulus had a duration of 0.67 s and the two signals were synchronised in time. In the later experiments employing a translating visual ‘object’, the stimulus was a Gabor patch and a fixation point was used. This was needed to keep the motion trajectory in a constant path on the retina and to ensure it followed the same path as the audio movement. The Gaussian envelope was 2.8° wide at half height and the spatial frequency of the vertical carrier grating was 2 cyc/deg. Gabor contrast was varied to find threshold for detection its motion amongst the dynamic visual noise (random changes in pixel intensity at 30 Hz).

2.2. Procedure

Motion detection thresholds were first measured for unimodal movement. This involved systematically varying the proportion of motion signal strength in the measured modality to home in on the detection threshold. This was done using the adaptive algorithm known as Quest [36], which was slightly modified to fit a cumulative Gaussian instead of a Weibull function (the Gaussian model provides a slightly better fit and returns a more intuitive pair of parameters for the offset and width of the function). While one modality was being measured, motion strength in the other, unmeasured modality was set to zero (static auditory noise at 0° azimuth, or random visual noise with no coherent motion). Thus, there were both visual and auditory stimuli present in unimodal conditions, importantly for comparisons with the bimodal conditions, but only one modality contained movement. Leftward and rightward movements were randomly intermingled over trials. As thresholds were very similar for both motion directions, data for the two directions were combined to obtain the final estimate of unimodal motion detection threshold. Although this left the threshold estimate almost unchanged, it served to improve the estimate of the slope parameter, which is critical for measurements of bimodal movement.

With unimodal thresholds established, auditory and visual motion were combined so that the bimodal motion detection thresholds could be measured. Again, this was done using the Quest algorithm, which was initiated with auditory and visual motion strength set to their unimodal threshold levels. As Quest varied, the motion strength of the bimodal stimulus to home in on its threshold, it was important that the unimodal motion components remained yoked at an equal probability level of detection. The subjects' unimodal psychometric functions were used to maintain the two components of the bimodal stimulus subjectively equated in intensity (see Fig. 3). Two bimodal conditions were compared: correlated, in which both motion trajectories were perfectly concordant (both leftwards, or both rightwards), and anti-correlated, in which the movement in the two modalities followed opposed trajectories (one leftward, the other rightward).

In all experiments, two-interval, forced-choice designs were used. One interval contained movement, the other no movement. The observer's task was to identify the interval containing the motion. Identifying the direction of the movement, which was randomly leftward or rightward, was not required. In bimodal conditions, subjects were free to use auditory, visual, or bimodal cues to make this judgement. Auditory and visual motions were calibrated and adjusted so that their trajectories were spatiotemporally coincident. For each threshold measurement, at least five Quest staircases were run for each condition, the data from which were pooled and fit with a cumulative Gaussian using a maximum likelihood procedure. Motion detection thresholds were defined as the motion strength corresponding to a 0.75 likelihood of correct detection. Both of the authors and two naïve observers provided data for these experiments.

3. Results

3.1. Visual and auditory motion thresholds

We first measured separately the coherence thresholds for discriminating the direction of motion of a visual and of an auditory sound source. Both stimuli were "broad-band" and designed to be as similar as possible (see illustration in Fig. 1). The visual stimulus—a field of 200 dots in which a random subset was displaced either leftward or rightward to create a sensation of coherent motion—was chosen because

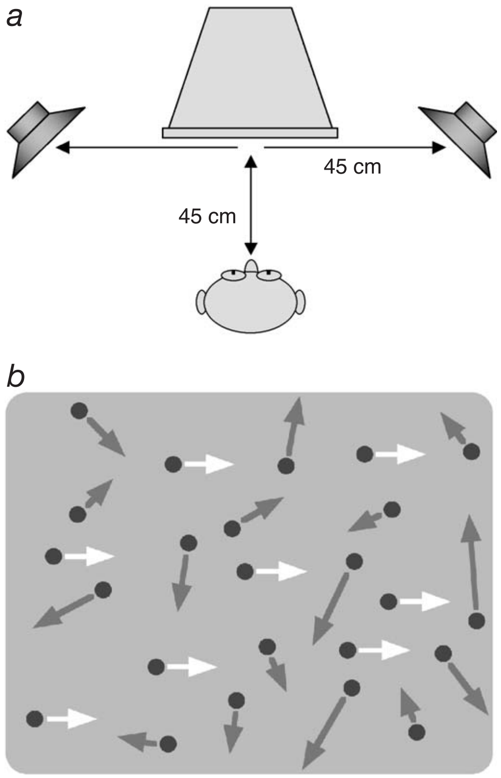

Fig. 1. Stimuli and apparatus: (a) Observers sat opposite a video monitor on which the visual stimuli were shown. Flanking the monitor and lying in the same plane as the screen were two speakers elevated to the screen's mid-height. Inter-aural time differences were used to create a sound source whose position moved smoothly from the left edge of the screen to the right edge. (b) The visual stimulus was composed of 200 small dots whose positions were updated at a rate of 50 Hz. A subset of the dots carried the motion signal (shown for illustrative purposes by white vectors) and were displaced uniformly to the left or right. The remainder of the dots (the noise dots, shown by black vectors) were replotted in random locations. Thus the noise dots could carry local motion signals in any direction or speed but collectively they contain no net movement.

it is known that variations in its motion coherence level elicit responses in visual cortical area MT which correlate highly with motion perception [2,3]. Two intervals of the visual stimulus were displayed, one containing purely incoherent random motion, the other containing a proportion of coherent motion mixed with random motion. Subjects identified which interval contained the movement (no judgement of direction was required). Over trials, the proportion of coherently moving dots was varied according to the observer's responses using an adaptive Quest routine [36] to determine the motion detection threshold. The results for four observers are shown in the upper panels of Fig. 2. Thresholds for leftward and rightward motion were very similar and did not differ statistically. The two data sets were therefore combined. Each graph plots the percentage of correct responses against a range of stimulus coherence levels. At moderate to high coherence levels, above 10%, performance was near perfect, dropping steadily to chance level (50%) at around 2% coherence. The curves were fit with a cumulative Gaussian to the pooled data set using a maximum-likelihood method. The threshold was defined as 75% correct performance, indicated by the dashed line. For all observers it was between 2% and 10%, consistent with previous research [20].

The auditory stimulus was designed to match as closely as possible the characteristics of the visual stimulus. It was thus composed of a motion signal (randomly leftward or rightward) and a noise component, and the two components varied in relative intensity to alter the strength of the motion signal. Again, the proportion of coherent motion was varied adaptively to home in on the threshold for auditory motion detection and data sets for leftward and rightward motion were pooled. The results for the same four observers are shown in the lower panels of Fig. 2. The pattern of results is similar to that for the visual task, except that the thresholds are generally higher.

In order to study any interaction between vision and audition, it is necessary to equate for the differences in absolute threshold. This is readily achieved by normalising the values of each to their measured unimodal thresholds. In addition, it may also be necessary to use the slopes of the unimodal psychometric functions to scale the stimuli, so that deviations from threshold are equated for probability of correct detection. Fortunately, under the conditions of this experiment, the slopes of the psychometric functions, defined as the inverse of standard deviation of the cumulative Gaussian, were very similar for the two modalities. As Fig. 2 shows, in both cases the standard deviations were very close to 0.3 log-units, with no significant trend for the auditory measures to differ from the visual. For this reason, the components of the bimodal stimuli were scaled solely in terms of threshold.

3.2. Bimodal motion thresholds

In the bimodal conditions, auditory and visual components were initially set to their individual thresholds and a Quest procedure was again used to search for the threshold of (bimodal) motion detection. As the strength of the bimodal stimulus varied from trial to trial, the individual visual and auditory components were kept yoked together in perceptual salience by scaling them as equal multiples of

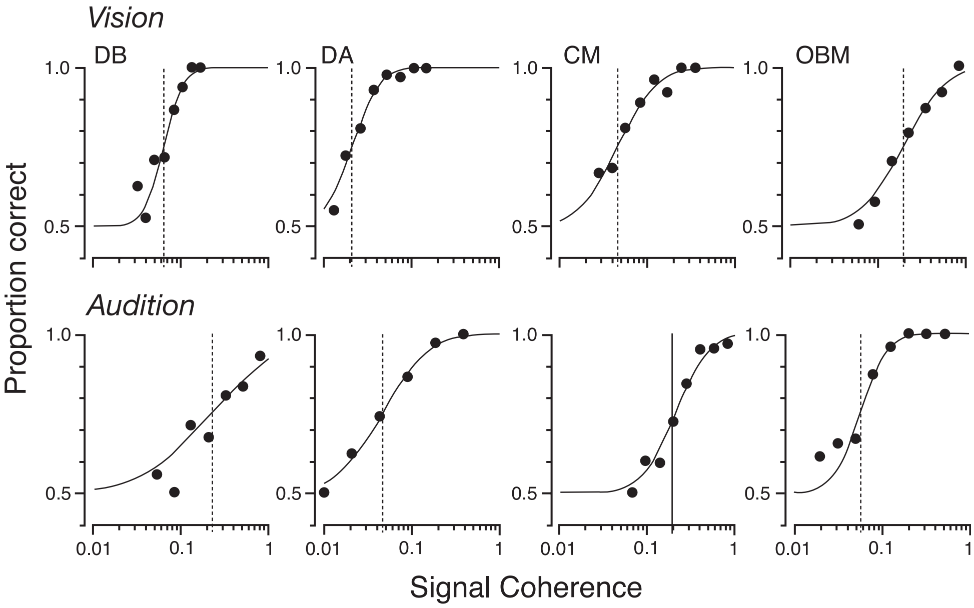

Fig. 2. Unimodal data: Probability of correct motion detection for the four subjects in the visual (upper) and auditory (lower) conditions. In both cases, the signal coherence was varied by an adaptive routine to home in on threshold. Threshold was taken as the 75% correct point of a cumulative normal function fitted to the data. For each subject, there were no statistical differences between the motion thresholds for leftward and rightward motion. This was true for motion in each modality. For reasons of reliability, the leftward and rightward data sets were pooled and a cumulative Gaussian was re-fit to the global data set, as shown above.

their individual thresholds. There were two conditions: a correlated condition, where the auditory and visual motions had the same speed and direction, and an anticorrelated conditions where the motion components had the same speed but their directions were opposed. As before, observers were simply required to identify which interval contained the coherent motion signal (and not its direction or whether or not it was correlated). The results are shown in Fig. 3, plotting percent correct against relative signal strength (normalised to the separate unimodal thresholds). In all cases, motion thresholds in the correlated bimodal conditions are less than one, indicating that bimodal performance was better than unimodal. However, the bimodal advantage was also obtained in the anticorrelated condition, where the two component motions moved in opposite directions. This is brought out most clearly in Fig. 4, which shows the bimodal thresholds for all observers for ‘same’ and for ‘opposite’ motion directions. Both conditions yielded very similar results with none of the observers exhibiting a significant difference between same or opposed motion directions. Averaging the thresholds for the four observers, the means for the same-direction condition (0.83) and opposite-direction condition (0.84) were virtually identical. Clearly, then, bimodal motion detection gains no advantage from the unimodal component motions having the same direction.

To understand the results better, the auditory coherence thresholds were plotted against visual coherence thresholds for the four observers (Fig. 5). For clarity, the data for leftward and rightward moving stimuli have been averaged so the plot is symmetrical. By definition, the unimodal thresholds for both vision and audition (cardinal axes) are

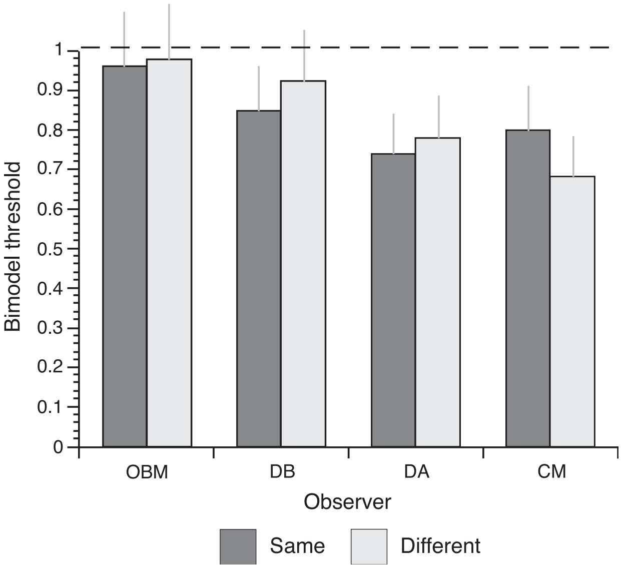

Fig. 4. Bimodal facilitation: Improvement in sensitivity to bimodal motion detection, relative to normalised unimodal thresholds (shown by the dashed line). Shaded bars show bimodal improvement when the auditory and visual components moved in opposite directions, and the open bars for movement in the same direction. Both conditions yielded very similar results (mean ‘same’ = 0.83; mean ‘opposite’ = 0.84) with none of the observers exhibiting a significant difference.

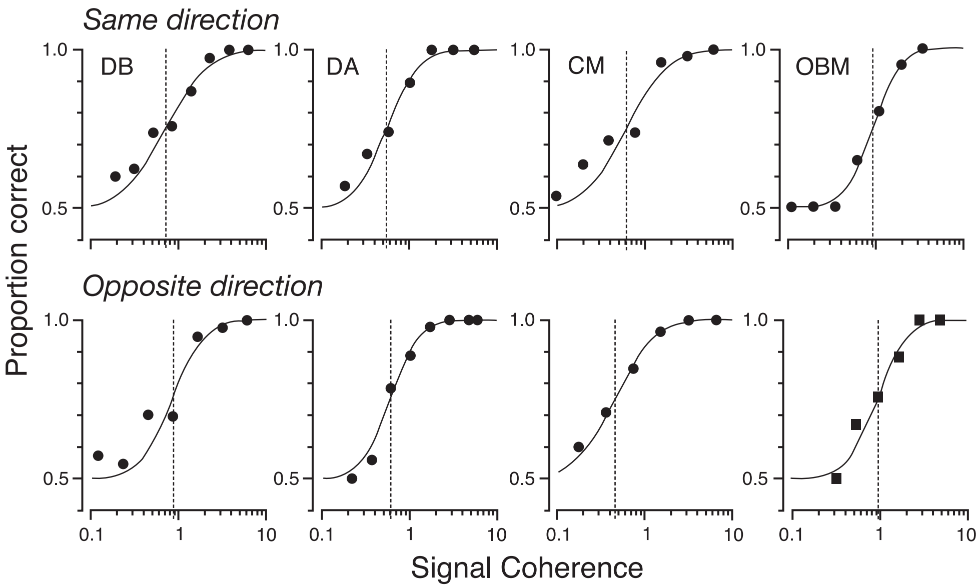

Fig. 3. Bimodal data: Probability of correctly detecting the interval containing coherent motion (visual and auditory together) for the four subjects. For the upper curves, the auditory and visual motions moved in the same direction (psychometric function is fit to the pooled LL and RR conditions), for the lower curves in opposite directions (psychometric function is fit to the pooled LR and RL conditions). The unimodal components were normalised to their own thresholds (from Fig. 1), so that a bimodal signal coherence of 1 indicates that each unimodal component was at threshold. As bimodal signal strength was varied to find the bimodal threshold, the unimodal component strengths were varied in equal multiples of threshold. This effectively yokes them together so that each component has the same probability of detection. The data show that the bimodal advantage is only slight, and is not consistently different for motion in the same versus different directions.

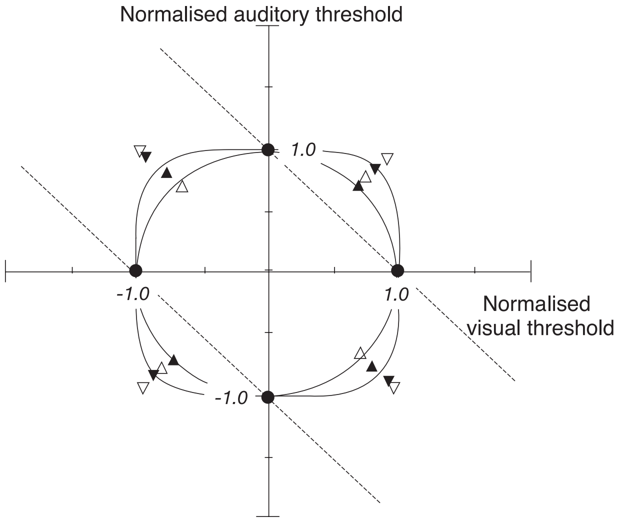

Fig. 5. Predicted audiovisual interactions based on three different models of cue combination. The measured unimodal thresholds have been normalized to unity and appear on the vertical and horizontal axes. The measured bimodal thresholds are shown by the triangular symbols lying on the oblique axes. The top right and lower left quadrants represent auditory and visual motion in the same directions (the LL and RR conditions differed little; they were averaged and are duplicated in each quadrant), and the top left and lower right quadrants represent motion in opposite directions (LR and RL also averaged and duplicated). The symbols are: open-upright triangles CM, open-inverted OBM, filled upright DA and filled-inverted DB. The broken curves indicate various predictions. The dashed straight lines indicate linear summation, producing very low thresholds for same direction motion, but high (infinite) thresholds when the direction is opposed. The dotted circle and dashed curves represent two simple models of non-linear summation. The dotted circle shows a maximum likelihood estimation model that combines optimally signals from the two modalities, based on the model of Ernst and Banks [7,10]. The motion vectors are squared (and hence lose their sign) before summation. The dashed curve shows a form of "probability summation", based on the steepness of the psychometric functions (Fig. 2). The thresholds are raised to the fourth power before summation [34], reflecting the slight increase in probability that one of the other motion type will reach threshold when both are displayed together. The data straddle the two non-linear summation predictions, and fall far from the linear summation prediction.

unity for all observers, as each was normalized by its own threshold. In the bimodal conditions (oblique axes), auditory and visual thresholds were identical for each observer (as motion strength was yoked) and the threshold is always less than unity.

To put this apparent improvement in context, three different predictions have been plotted. The dashed oblique lines show predictions for linear summation: at all points along this curve, the auditory and visual signals sum to unity. As the sum is signed, the curves tend to infinity when the directions are opposite. This model assumes that the motion signals for auditory and visual movement are summed (linearly) before the decision about the presence or absence of bimodal movement is made. The continuous curve resembling a rounded square shows another form of combination of auditory and visual signals, termed "probability summation". On this model, the two motion signals are processed independently. Probability summation would occur if there were no interaction between the auditory and visual signals at the processing stage, but the observer was monitoring both of them at a decision stage. If either signal (or both) reached threshold, the observer would use that information for his or her judgement. This leads to a slight "probabilistic" improvement in performance (see Section 2 for details) for bimodal stimuli since the likelihood of detecting at least one stimulus is greater when two are presented. Critically, however, the improvement is unsigned: identical threshold improvements are obtained irrespective of whether the component motions are the same or are opposed. A final model is shown by the dotted circle. It indicates the predicted improvement based on ideal statistical combination of information across senses, similar to the maximum likelihood estimation model proposed by Ernst and Banks [7]. This model predicts a "Pythagorean" improvement in thresholds in the bimodal condition, indicated by the circle. Again, if the assumptions of the model are to combine an unsigned signal of motion, the predictions will be symmetrical for correlated and anti-correlated bimodal motion.

The pattern of results is clearly not consistent with linear summation of signed motion signals: the level of summation observed is too small for this, and it does not show the asymmetry towards like-direction that would be expected. The summation is however consistent with a statistical combination of signals, either simply by probability summation (increased chance of a response by sampling more independent detectors), or by a statistical combination based on maximum likelihood estimation [7,10]. The predictions of both classes of model are too close for our data to decide between.

3.3. Motion thresholds for discrete objects in central and peripheral space

It might be argued that the absence of an audiovisual interaction resulted from the visual and auditory motion signals not being appropriate for bimodal integration. That is, while the auditory motion stimulus could be interpreted as a single, translating object (being a diffuse point-source), the visual stimulus could not, as the motion was spatially distributed over the entire screen and intermingled with noise elements. Conceivably, the perceptual system might veto the combination of auditory and visual motion signals if they violate ecological constraints. To address this, we conducted an experiment wherein we measured visual motion detection thresholds for a translating visual object rather than for random dot motion. Only visual motion thresholds were measured, and as above, this was done in the presence of accompanying auditory motion signals with the same or the opposite direction. In this way, when the auditory and visual motions were correlated, both motion signals were consistent with a coherently translating single object.

In addition to these bimodal motion conditions, two control conditions were included: one in which visual motion detection was measured with no acoustic stimulus present, and another in which the auditory stimulus was present but stationary (at zero degrees azimuth). In conditions involving an auditory stimulus (whether stationary or translating), the sound was salient and suprathreshold at 72 dB A. In the case where the auditory signal was translating, it contained 100% motion signal. The visual stimulus was a sinusoidal grating windowed by a Gaussian envelope (a Gabor patch) whose width was 2.8° at half height and which translated across the video monitor from left to right or from right to left. To measure the detection threshold for the visual stimulus, the translating Gabor patch was masked by dynamic visual noise. The contrast of the Gabor was varied adaptively (Quest) to find the threshold for visual motion detection. Results for this experiment are shown in Fig. 6b (see 'central' vision). For both observers, the amount of visual noise eliciting threshold-level performance in detecting the visual motion was essentially identical and independent of whether the accompanying auditory movement was correlated or was anticorrelated (two-tailed *t*-tests for both observers were not significant at the 5% level). This outcome shows a lack of audiovisual interaction for motion detection and therefore supports our initial result. Further support comes from our finding that motion detection thresholds in the two control conditions (sound present but stationary or sound absent) were essentially the same as in the two experimental conditions involving auditory motion.

The absence of an audiovisual interaction might also be due to the fact that the visual motion signal was presented in the central visual field. The visual system has excellent resolution in central vision and would gain little from incorporating acoustic motion signals. However, visual performance in the peripheral visual field is poorer and might be improved by incorporating an auditory motion signal that is spatiotemporally correlated with the visual motion. Very recent findings showing direct connections between primary auditory cortex and the peripheral representation in primary visual cortex give credence to such a possibility [19]. To explore this possibility, we repeated the experiment with the visual motion trajectory displaced by 25° into the upper or lower periphery.

In order to examine audiovisual integration in the visual periphery, the visual fixation point (rather than the visual motion) was relocated to a peripheral location

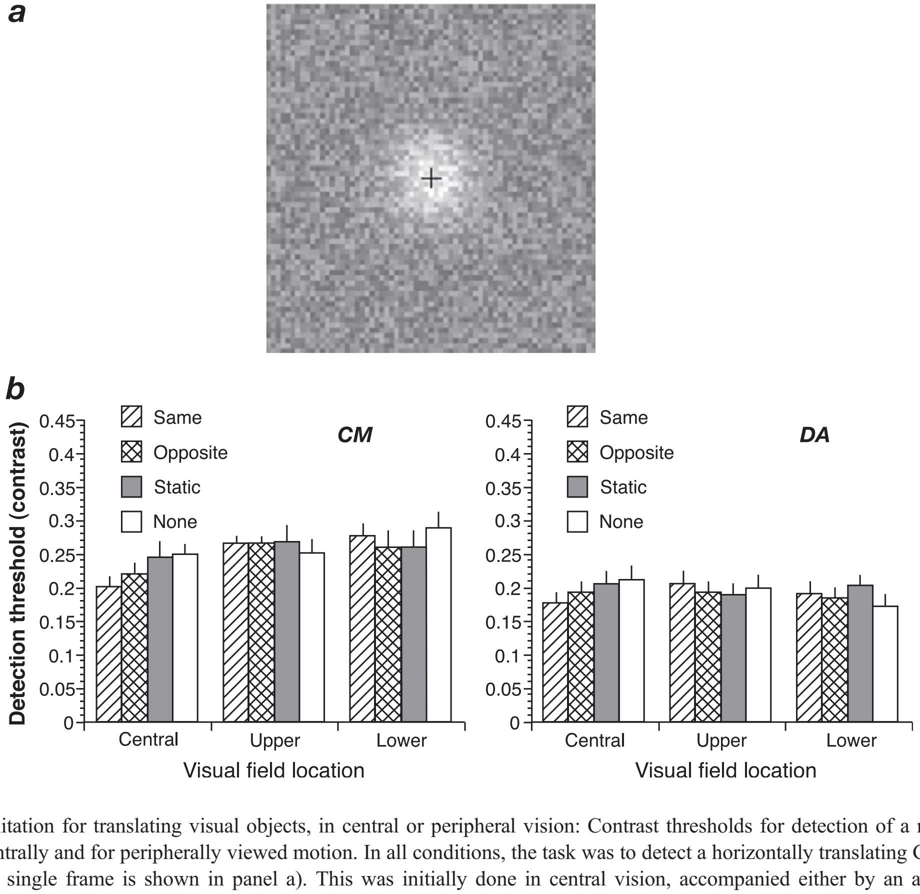

Fig. 6. No bimodal facilitation for translating visual objects, in central or peripheral vision: Contrast thresholds for detection of a moving visual object are shown in panel b, for centrally and for peripherally viewed motion. In all conditions, the task was to detect a horizontally translating Gabor patch embedded in dynamic visual noise (a single frame is shown in panel a). This was initially done in central vision, accompanied either by an auditory motion that was correlated (striped columns) or anticorrelated (crossed columns) with the visual motion, or by a stationary auditory signal (gray column). There were no significant differences among these bimodal conditions, and they did not differ from the unimodal visual condition (white column). These results argue strongly against any audiovisual interaction in the detection of translating bimodal stimuli. The same pattern of results obtained when the experiment was repeated 25° into the upper or lower visual field, despite a large degradation of visual acuity in the peripheral visual field. For peripheral presentation of the visual stimulus, the fixation point was vertically displaced, therefore leaving the visual and auditory motion paths superimposed.

above or below the visual motion trajectory. This ensures the visual and auditory trajectories remained spatially superimposed (as in the centrally viewed conditions) and that differences in performance cannot be attributed to artifacts in the video monitor. The pattern of results (Fig. 6b, see 'upper' and 'lower' periphery) across the four conditions is essentially identical to that obtained in central vision. Clearly, then, detection thresholds for peripheral visual motion are unaffected by the absence or presence of an acoustic stimulus, irrespective of whether that acoustic stimulus be static or translating—correlated or anticorrelated—with respect to the translating visual object. Thus, our original conclusion of a lack of audiovisual interaction for motion detection is again supported and can be extended to encompass the peripheral visual field as well as central vision. This latter extension may seem surprising since it is known that visual acuity is generally worse in the periphery. Indeed, spatial and temporal contrast sensitivity both decrease linearly with eccentricity, although the attenuation for temporal acuity is shallower [34,38]. Thus it appears that visual and auditory motion signals are not integrated in the periphery even though it would be advantageous to do so.

4. Discussion

Taken together, these results show a small non-directional gain in bimodal movement detection for bimodal motion, but contain no evidence of a facilitative audiovisual interaction. This holds true for both coherently moving visual objects and for spatially distributed motions, in central and in peripheral vision. These conclusions agree with two recent reports [18,39]. As in our study, both used spatially distributed random-dots to create visual motion that was either weakly visible or subthreshold. This was paired with a suprathreshold auditory movement produced by cross-fading two loudspeakers. In the first paper [18], subjects were presented with a single-interval trial and judged whether the direction of visual movement was leftward or rightward. They report that when visual movement was subthreshold, subjects' judgements of visual direction were biased to match the direction of auditory movement (which was fixed in intensity and clearly audible at all times). Apart from inducing a response bias, and consistent with our findings, auditory movement was found not to affect sensitivity to visual motion. In a second paper, Wuerger et al. [39] independently conducted an experiment using a design virtually identical to our first experiment to study how auditory and visual motion signals interact at threshold. Their finding, like ours, was that there are no threshold interactions beyond those predicted by simple probability summation, with the individual motion signals extracted independently and then combined at a decision stage.

While these two studies are similar in design, they differ slightly in important details. Wuerger et al. used interaural intensity differences in stereo white noise to generate auditory movement, whereas we used interaural timing differences in low-pass noise. This is significant in that timing-based and intensity-based sound localization is carried out by separate auditory processes. Other differences are that their auditory and visual stimuli spanned relatively small spatial ranges (7.4° and 16°, respectively) while ours spanned larger spatial ranges (40° and 50°, respectively), and that our stimulus durations were 670 ms, whereas theirs were 175 ms for vision and 1000 ms for audition. One other difference between the studies concerns the way in which auditory movement was defined. In our study, auditory movement was 'apparent' in that it was defined by a rapid sequence of static locations (41 locations in 0.67 s, the same 60 Hz update rate as in our visual stimulus), whereas the movement in the Wuerger et al. study was a continuous cross-fading of stereo speaker levels. While it might be argued that apparent motion is not ecologically equivalent to smooth auditory movement, it is also the case that smooth cross-fading has its limitations. For example, it changes the speaker levels so that the direction of motion can be determined by listening to the level of just one speaker. Also, it can become perceptually bistable after repeated listening and be heard as either two independent loudspeakers whose levels rise and fall in antiphase or as auditory movement. The stimuli in our study provided a compelling percept of continuous auditory movement. In any event, stimulus differences appear not to have been critical as the similar results and conclusions demonstrate. In finding a similar pattern of data, despite stimulus differences, these studies indicate that the lack of a facilitative interaction between auditory and visual motion at threshold is a general result.

One aspect of our paper which is distinct from Wuerger et al. concerns our final experiment in which a translating visual object was used instead of a distributed visual signal. One limitation of the RDK stimulus is that it has no clear spatial focus, in contrast to the auditory motion stimulus which had a clearly defined focus. Our results for this experiment show that the absence of a co-localised motion stimulus was not the reason for the lack of facilatory threshold interactions observed in the first bimodal experiment using the RDK. One potentially important factor in this lack of interaction concerns the auditory stimulus. Movement of a real-world sound source would entail changes in the three cues to auditory localisation: interaural timing and intensity differences, and monaural spectral cues. We produced auditory movement by varying only interaural timing differences, and Wuerger et al. varied only interaural intensity. Perhaps the use of a more ecological valid auditory stimulus would reveal greater interaction with visual movement? Answering this question is possible. It would require, for example, the use of a robotic arm programmed to translate a loudspeaker smoothly, or perhaps a finely spaced alignment of matched speakers. In future work, we will use virtual auditory space technology [4] to investigate this question.

Clear evidence of audiovisual interactions has been found in other areas of psychophysical study. Perhaps the most well known instance of this is the McGurk effect [17] in which a single phoneme is misidentified if the listener watches a human speaker mouthing a conflicting phoneme (classically, an auditory/ba/paired with a visual/ga/often produces the percept/da/). In another early study, the rate of a fluttering auditory stimulus was found to influence the rate of flicker perceived in a modulating visual stimulus [24]. This latter study is similar in nature to a recent report [22] which showed that a briefly flashed visual stimulus presented in the near periphery can be perceived to flash twice if it is accompanied by two short auditory beeps. Both studies demonstrate a cross-modal influence of audition on vision, in contrast to the McGurk effect, which reveals a cross-modal influence operating in the other direction. However, the influence of audition on vision is not limited to speech. Sound has also been used to resolve an ambiguous visual motion stimulus in which two discs oscillate in antiphase along the same horizontal trajectory [21]. The stimulus is bistable: it may be seen as two discs rebounding off each other (and reversing direction), or as two discs passing through each other (maintaining their direction). The tendency to see the discs rebounding instead of passing through each other was strengthened by an auditory 'click' at the moment of 'impact', illustrating that a simple sound cue can be used to disambiguate an underspecified visual stimulus. All of these studies are related in demonstrating that the phenomenology of a stimulus in one sensory modality can be altered by signals in another. Moreover, this influence appears to be early: an evoked potential study indicates that Shams et al.'s [23] "double flash" illusion modulates early components of the evoked potential over the occipital lobe.

The present study can be distinguished from those reviewed above. We did not seek to alter the phenomenology of a stimulus in one modality by the presence of a second hetero-modality stimulus. Rather, our experiment examined whether the sensitivity of the perceptual system to a particular attribute (motion, in this case) is improved when that attribute is presented simultaneously to two modalities. We find that there is no improvement beyond what is expected on the basis of statistical combination of signals, either a maximum likelihood estimation [7] or probability summation [33]. Our finding that stimuli containing opposed motion directions yield the same bimodal improvement as stimuli containing the same motion directions is very strong evidence for this type of model. Of course, these results were obtained for horizontal trajectories in the fronto-parallel plane, and our conclusion, strictly, is limited to this case. It remains an open question whether other forms of motion might benefit from an audiovisual integration or correlated motion signals and continuing work in our laboratory is investigating this possibility.

Acknowledgements

David Alais is supported by a Marie Curie Fellowship from the European Commission.

References

[1] C. Auerbach, P. Sperling, An auditory–visual space: evidence for its reality, Percept. Psychophys. 16 (1974) 129–135.

[2] K.H. Britten, M.N. Shadlen, W.T. Newsome, J.A. Movshon, The analysis of visual motion: a comparison of neuronal and psychophysical performance, J. Neurosci. 12 (1992) 4745–4765.

[3] K.H. Britten, W.T. Newsome, M.N. Shadlen, S. Celebrini, J.A. Movshon, A relationship between behavioral choice and the visual responses of neurons in macaque MT, Vis. Neurosci. 13 (1996) 87–100.

[4] S. Carlile, Virtual Auditory Space: Generation and Applications, Chapman & Hall, New York, 1996.

[5] C.L. Colby, J.-R. Duhamel, M.E. Goldberg, Ventral intraparietal area of the macaque: anatomic location and visual response properties, J. Neurophysiol. 69 (1993) 902–914.

[6] R. Desimone, M. Wessinger, L. Thomas, W. Schneider, Attentional control of visual perception: cortical and subcortical mechanisms, Cold Spring Harbor Symp. 55 (1990) 963–971.

[7] M.O. Ernst, M.S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature 415 (2002) 429–433.

[8] M.A. Frens, A.J. Van Opstal, R.F. Van der Willigen, Spatial and temporal factors determine audio–visual interactions in human saccadic eye movements, Percept. Psychophys. 57 (1995) 802–816.

[9] M.S. Graziano, A system of multimodal areas in the primate brain, Neuron 29 (2001) 4–6.

[10] J.M. Hillis, M.O. Ernst, M.S. Banks, M.S. Landy, Combining sensory information: mandatory fusion within but not between the senses, Science 298 (2002) 1627–1630.

[11] D. Kadunce, J. Vaughan, M. Wallace, G. Bendek, B. Stein, Mechanisms of within- and cross-modality suppression in superior colliculus, J. Neurophysiol. 78 (1997) 2834–2847.

[12] A.J. King, A.R. Palmer, Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus, Exp. Brain Res. 60 (1985) 492–500.

[13] J. Lewald, R. Guski, Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli, Brain Res. Cogn. Brain Res. 16 (2003) 468–478.

[14] J. Lewis, D. Van Essen, Corticocortical connections of visual, sensorimotor and multimodal processing areas in parietal lobe of macaque monkey, J. Comp. Neurol. 428 (2000) 112–137.

[15] M. Meredith, B.E. Stein, Spatial determinants of multisensory integration in cat superior colliculus, J. Neurophysiol. 75 (1996) 1843–1857.

[16] M. Meredith, J.W. Nemitz, B.E. Stein, Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors, J. Neurosci. 7 (1987) 3215–3229.

[17] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.

[18] G.F. Meyer, S.M. Wuerger, Cross-modal integration of auditory and visual motion signals, NeuroReport 12 (2001) 2557–2560.

[19] K.S. Rockland, H. Ojima, Calcarine area V1 as a multimodal convergence area, Abstr.-Soc. Neurosci. 27 (2001) 511.20.

[20] M.O. Scase, O.J. Braddick, J.E. Raymond, What is noise for the motion system? Vis. Res. 36 (1996) 2579–2586.

[21] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, Nature 385 (1997) 308.

[22] L. Shams, Y. Kamitani, S. Shimojo, Illusions: what you see is what you hear, Nature 408 (2000) 788.

[23] L. Shams, Y. Kamitani, S. Thompson, S. Shimojo, Sound alters visual evoked potentials in humans, NeuroReport 12 (2001) 3849–3852.

[24] T. Shipley, Auditory flutter-driving of visual flicker, Science 145 (1964) 1328–1330.

[25] R.J. Snowden, O.J. Braddick, Differences in the processing of short-range apparent motion at small and large displacements, Vis. Res. 30 (1990) 1211–1222.

[26] D.L. Sparks, Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus, Physiol. Rev. 66 (1986) 118–171.

[27] C. Spence, J. Driver, Audiovisual links in endogenous covert spatial attention, J. Exp. Psychol. Hum. Percept. Perform. 22 (1996) 1005–1030.

[28] C. Spence, J. Driver, On measuring selective attention to a specific sensory modality, Percept. Psychophys. 59 (1997) 389–403.

[29] G.P. Standage, L.A. Benevento, The organization of connections between the pulvinar and visual area MT in the macaque monkey, Brain Res. 262 (1983) 288–294.

[30] B.E. Stein, Neural mechanisms for synthesizing sensory information and producing adaptive behaviors, Exp. Brain Res. 123 (1998) 124–135.

[31] B.E. Stein, W.S. Huneycutt, M.A. Meredith, Neurons and behavior: the same rules of multisensory integration apply, Brain Res. 448 (1988) 355–358.

[32] B.E. Stein, M. Meredith, W. Huneycutt, L. McDade, Behavioral indices of multisensory integration: orienting to visual cues is affected by auditory stimuli, J. Cogn. Neurosci. 1 (1989) 12–24.

[33] C.W. Tyler, C.C. Chen, Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation, Vis. Res. 40 (2000) 3121–3144.

[34] V. Virsu, J. Rovamo, P. Laurinen, Temporal contrast sensitivity and cortical magnification, Vis. Res. 22 (1982) 1211–1217.

[35] M. Wallace, M. Meredith, B.E. Stein, Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus, J. Neurophysiol. 69 (1993) 1797–1809.

[36] A.B. Watson, D.G. Pelli, QUEST: a Bayesian adaptive psychometric method, Percept. Psychophys. 33 (1983) 113–120.

[37] L.K. Wilkinson, M.A. Meredith, B.E. Stein, The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior, Exp. Brain Res. 112 (1996) 1–10.

[38] M.J. Wright, A. Johnston, Spatiotemporal contrast sensitivity and visual field locus, Vis. Res. 23 (1983) 983–989.

[39] S.M. Wuerger, M. Hofbauer, G.F Meyer, The integration of auditory and visual motion signals at threshold, Percept. Psychophys., 2004 (in press).