EJN fens EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, Vol. 31, pp. 1763–1771, 2010 doi:10.1111/j.1460-9568.2010.07198.x

Acquired prior knowledge modulates audiovisual integration

Marc M. Van Wanrooij, Peter Bremen and A. John Van Opstal
Radboud University Nijmegen, Donders Institute of Brain, Cognition and Behaviour, Department of Biophysics, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

Keywords: human, multisensory integration, orienting, saccades, spatial behaviour

Abstract

Orienting responses to audiovisual events in the environment can benefit markedly by the integration of visual and auditory spatial information. However, logically, audiovisual integration would only be considered successful for stimuli that are spatially and temporally aligned, as these would be emitted by a single object in space–time. As humans do not have prior knowledge about whether novel auditory and visual events do indeed emanate from the same object, such information needs to be extracted from a variety of sources. For example, expectation about alignment or misalignment could modulate the strength of multisensory integration. If evidence from previous trials would repeatedly favour aligned audiovisual inputs, the internal state might also assume alignment for the next trial, and hence react to a new audiovisual event as if it were aligned. To test for such a strategy, subjects oriented a head-fixed pointer as fast as possible to a visual flash that was consistently paired, though not always spatially aligned, with a co-occurring broadband sound. We varied the probability of audiovisual alignment between experiments. Reaction times were consistently lower in blocks containing only aligned audiovisual stimuli than in blocks also containing pseudorandomly presented spatially disparate stimuli. Results demonstrate dynamic updating of the subject's prior expectation of audiovisual congruency. We discuss a model of prior probability estimation to explain the results.

Introduction

Each of our senses extracts information about events in the environment. Successful integration of these separate information streams can be highly beneficial in numerous tasks, ranging from improved stimulus identification to speeding of orienting responses and enhanced localisation performance. In the present study we focus on the control of rapid head saccades toward a novel audiovisual stimulus in the peripheral visual field.

A large body of experimental evidence has indicated that audiovisual integration leads to a marked reduction in saccadic reaction times for co-occurring and spatially aligned audiovisual targets. Typically, experiments tested relatively simple conditions in which a single auditory and a single visual stimulus could occupy a limited number of possible configurations (Hughes *et al.*, 1994; Nozawa *et al.*, 1994; Frens *et al.*, 1995; Goldring *et al.*, 1996). However, for more complex audiovisual scenes that contain more uncertainty about upcoming target locations and audiovisual combinations, perceptually aligned audiovisual events also consistently produce faster and more accurate orienting responses than their unisensory counterparts (Corneil *et al.*, 2002; Van Wanrooij *et al.*, 2009).

The mechanisms and rules that govern audiovisual integration to evoke rapid and accurate orienting behaviour have been explained by neural interactions within spatially organized neural maps, such as in the midbrain superior colliculus (Stein & Meredith, 1993; Frens & Van Opstal, 1998; Wallace *et al.*, 1998; Bell *et al.*, 2005). These interactions induce excitatory effects for congruent multisensory inputs but suppress each other when they fall outside the spatial–temporal integration window (Meredith & Stein, 1986a; Meredith *et al.*, 1987). A theoretical account of such a mechanism was offered by Anastasio *et al.* (2000), who proposed that the principle of optimal statistical inference (Bayesian reasoning) underlies multisensory integration.

Strict integration of audiovisual cues (Alais & Burr, 2004), however, is not always the most desirable option (Hillis *et al.*, 2002). In such a case, audiovisual events emanating from distinct objects would also be integrated, thus losing their segregation and identities. Obviously, the brain is able to cope with such situations and can readily distinguish spatially disparate audiovisual stimuli (Wallace *et al.*, 2004; Kording *et al.*, 2007; Sato *et al.*, 2007). The breakdown of multisensory integration is also demonstrated by a systematic reduction in speed and accuracy of saccadic eye movements to disparate stimuli (Frens *et al.*, 1995; Harrington & Peck, 1998; Hughes *et al.*, 1998; Colonius & Arndt, 2001; Van Wanrooij *et al.*, 2009).

How does the brain know which auditory and visual signals to fuse into an integrated percept, and which not? According to Bayesian models of multisensory integration the strength of multisensory fusion is proportional to the amount of coupling between sensory streams, which reflects the prior knowledge that multisensory inputs belong together (Ernst, 2005). Such prior knowledge could be based on experience and is likely to be adaptive.

Correspondence: Dr Marc M. Van Wanrooij, as above.
E-mail: m.vanwanrooij@donders.ru.nl

Received 4 October 2009, revised 1 December 2009, accepted 11 December 2009

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Here we address the question whether subjects extract and use information about expected alignment of audiovisual stimuli on the basis of the experimental stimulus statistics, and accordingly adapt multisensory integration when orienting to audiovisual stimuli. To that end, we manipulated the proportion of audiovisual spatial congruency in separate experimental blocks. If subjects dynamically adjust their expectation of audiovisual congruency, one expects enhanced integration effects when the probability of spatial alignment is high and decreased integration for low probabilities. Our results corroborate this hypothesis.

Materials and methods

Subjects

Seven subjects, aged 21–33 (mean, 27.7 years), participated in this study. Two subjects (MW and PB) are authors of this paper; the remaining five participants were naive about the purpose of the study. All subjects had normal hearing (within 20 dB of audiometric zero) as determined by an audiogram obtained with a standard staircase procedure (10 tone pips, 0.5-octave separation, between 500 Hz and 11.3 kHz) and had normal or corrected (MA, MW) binocular vision, with the exception of subject PB who did not wear his prescription glasses during the experiments. As the flash was supra-threshold, and his V and AV responses were within the normal range, we included his results in this study.

Experiments were conducted after subjects gave their full understanding and written consent. The experimental procedures were approved by the Local Ethics Committee of the Radboud University Nijmegen and adhered to The Code of Ethics of the World Medical Association (Declaration of Helsinki), as printed in the British Medical Journal of July 18, 1964.

Apparatus

During the experiments, subjects sat comfortably in a chair in the centre of a completely dark, sound-attenuated room (3 × 3 × 3 m). The floor, ceiling and walls were covered with sound-attenuating black foam (50 mm thick with 30 mm pyramids; AX2250, Uxem b.v., Lelystad, The Netherlands), effectively eliminating echoes for frequencies exceeding 500 Hz. The room had an ambient background noise level of ~30 dB SPL.

The chair was positioned at the centre of a vertically oriented circular hoop (radius 1.2 m) on which an array of 29 small broad-range loudspeakers (SC5.9; Visaton GmbH, Haan, Germany) was mounted at 5° intervals from −55 to +85° in the midsagittal plane (elevation angles, with 0° at straight ahead). Acoustic stimuli were digitally generated using Tucker-Davis System 3 hardware (Tucker-Davis Technologies, Alachua, FL, USA), with a real-time processor (RP2.1 System3, 48,828-Hz sampling rate). All acoustic stimuli consisted of 65-dB (A-weighted), 50-ms Gaussian white noise (0.5–20 kHz bandwidth), with 0.5 ms sine-squared onset and cosine-squared offset ramps. Visual stimuli consisted of green (wavelength 565 nm) light-emitting diodes (LEDs) mounted at the centre of each speaker (luminance 0.5 cd/m^2).

Head movements were recorded with the magnetic search-coil technique (Robinson, 1963). To this end, the listener wore a lightweight spectacle frame with a small coil attached to its nose bridge. Three orthogonal pairs of square coils (6 mm^2 wires, 3 m × 3 m) were attached to the room's edges to generate the horizontal (80 kHz), vertical (60 kHz) and frontal (48 kHz) magnetic fields, respectively. The head-coil signal was amplified and demodulated (EM7; Remmel Labs, Katy, TX, USA), low-pass-filtered at 150 Hz (custom built, fourth-order Butterworth), and digitized by a Medusa Head Stage and Base Station (TDT3 RA16PA and RA16; Tucker-Davis Technology) at a rate of 1017.25 Hz per channel.

A custom-written C++ program running on a PC (Precision 380; 2.8 GHz Intel Pentium D; Dell, Limerick, Ireland) controlled data recording and stimulus generation.

Experiments

We performed six different experiments: a calibration experiment, a unisensory auditory (A) and visual (V) experiment, and three audiovisual (AV) experiments. The AV experiments differed mainly in their distributions of spatial disparities between auditory and visual stimuli, as outlined below. Apart from the calibration experiment, all stimuli in the V, A and AV experiments were presented in the midsagittal plane. Note that the auditory system has different mechanisms to localise sounds in azimuth and in elevation (binaural difference cues for azimuth, and spectral cues for elevation). It is therefore not trivial how two-dimensional spatial disparity influences audiovisual integration (see also Corneil *et al.*, 2002; Van Wanrooij *et al.*, 2009). However, we have opted for the current study to utilise only vertical disparity, in order to focus on the effect of expectation.

Every experimental session began with the visual calibration experiment, followed by four blocks of a single other experiment. Each session was performed on a separate day.

In all experiments, except for the calibration experiment, subjects initiated a trial by a button press after first fixating a straight-ahead fixation LED. This button press extinguished the fixation LED 100–200 ms later; this was immediately followed by the visual target flash (50 ms) and/or a synchronous sound. Subjects were instructed to direct a head-fixed laser pointer (attached to the spectacle frame required for head movement recording; see Apparatus section) as quickly and as accurately as possible to the target location. In the AV experiments, the target was always the visual flash. As reaction times were typically > 100 ms, all responses were made under open-loop conditions.

Calibration experiment

To obtain the head-position data for the calibration procedure subjects accurately pointed the head-fixed laser pointer towards 56 LED locations in the two-dimensional frontal hemifield that encompassed the stimulus range of the actual experiments. Each experimental session started with this calibration run.

Visual experiment

Visual targets (50 ms duration) were presented at 10 possible locations: ± 15, 20, 25, 30 or 35° in elevation. Each location was presented 80 times, yielding 800 trials. These trials were pseudorandomly presented in four separate consecutive blocks of 200 trials. In between those blocks, subjects were allowed a short break (1–3 min) in which the room lights were illuminated.

Auditory experiment

The same target locations as in the visual experiment were employed, with targets being auditory.

Audiovisual 100%-aligned/0%-distractor experiment (AV-100/0). In the AV-100/0 experiment, the visual targets (same locations as in the

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

visual experiment) were accompanied by spatially and temporally aligned sounds. Subjects were specifically instructed that this was the case, and that the flash was at the target location.

Audiovisual 10%-aligned/90%-distractor experiment (AV-10/90). All spatial combinations of flash and sound locations were presented eight times in the AV-10/90 experiment, yielding 800 trials (10 flash locations × 10 sound locations × 8 repetitions). As with the visual experiment, these trials were pseudorandomly presented in four consecutive blocks of 200 trials. In this experiment, the sounds thus provided no *a priori* knowledge about the visual target. Subjects were instructed beforehand that the sound could be ignored. Only 10% of all trials contained spatially aligned audiovisual stimuli.

Audiovisual 50%-aligned/50%-distractor experiment (AV-50/50). In the AV-50/50 experiment, 50% of all trials contained spatially aligned AV stimuli while the other 50% had a spatial disparity > 45°. The 300 trials (10 flash locations × 2 disparities × 15 repetitions) in this experiment were divided into two blocks of 150 trials. Subjects were instructed, as in the AV-10/90-experiment, that the sounds could be ignored.

Distributions of spatial disparities

The different spatial distributions of each AV experiment are depicted in Fig. 1A–C. Due to the pseudorandom presentation of stimuli, each trial could be preceded by various combinations of disparate or congruent trials (disparate is defined as a spatial disparity > 45° and congruent as < 15°). This trial order, which is an important aspect of our experimental rationale, is exemplified in Fig. 1D; a congruent (C) trial might be preceded by either a congruent trial (CC), or a disparate (D) trial (DC). This preceding trial was also preceded by either a congruent or a disparate trial, leading to four different triplet trial sequences ending in a congruent trial. As an example, the distribution of uninterrupted series of either congruent or disparate trials is shown in Fig. 1E and F.

The spatial disparities in these experiments are expressed as physical disparities between the visual and auditory stimulus locations instead of perceived spatial disparities as described in Van Wanrooij *et al.* (2009). In the present study no background noise was added, and the auditory localization responses were highly accurate (slope of stimulus–response relations close to 1; data not shown). Therefore, physical and perceived spatial disparities were indistinguishable.

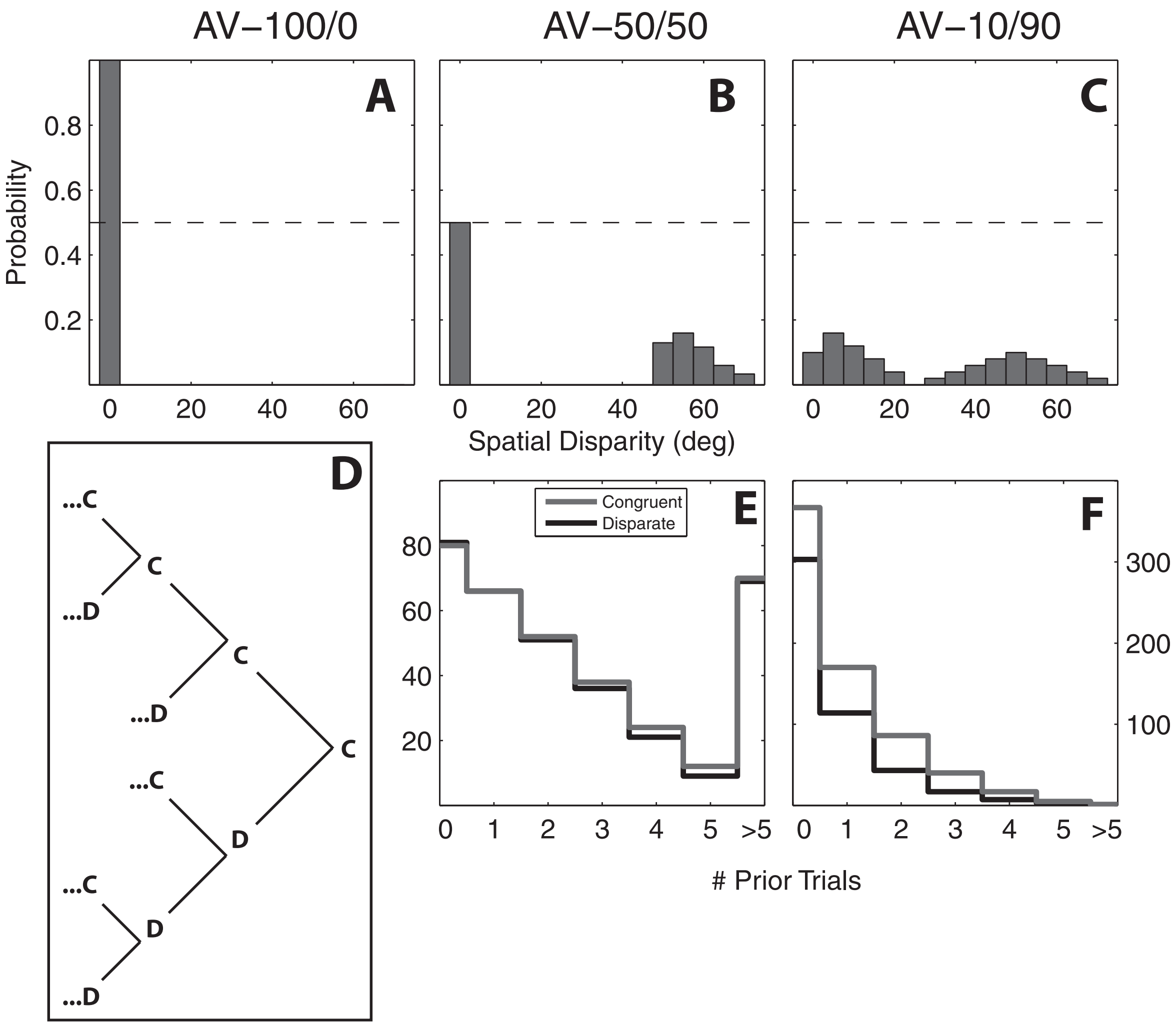

FIG. 1. Distribution of AV spatial disparities in the (A) AV-100/0, (B) AV-50/50 and (C) AV-10/90 experiments. (D) The pseudorandom nature of trial presentation led to different combinations of trial disparity order. The number of trials for uninterrupted series of congruent (disparity< 15°) and disparate (disparity> 45°) stimuli is shown for (E) the AV-50/50 experiment and (F) the AV-10/90 experiment. Grey line, congruent sequence; black line, disparate sequence.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Data analysis

All data analyses were performed off-line in MatLab (r2008a; The Mathworks, Natick, MA, USA).

Data calibration. Response data were calibrated by training two three-layer neural networks with the back-propagation algorithm that mapped final head orientations onto the known target positions of the visual calibration experiment (Goossens & Van Opstal, 1997). Head-position data from the other experiments were calibrated off-line using these networks with an absolute accuracy < 3% over the entire range. Head movements were automatically detected from calibrated data based on velocity criteria (onset > 20°/s, offset < 15°/s). Onset and offset markings were visually checked by the experimenter, and adjusted if necessary.

Performance. Performance of the subjects was quantified by the reaction time (onset head movement – onset target), and localisation error (see below) of the first goal-directed head movement in a trial. Responses with reaction times < 60 ms, or > 600 ms, or with an amplitude < 5°, were discarded from the analysis, as they were deemed to be due to prediction or to inattentiveness of the subject.

Localisation error. We quantified localisation accuracy by linear regression on the target-response relation:

$$\varepsilon_R = a \cdot \varepsilon_T + b \tag{1}$$

where ε_R and ε_T are response elevation and target elevation, respectively. Note that in all experiments except for the auditory-only experiment, the target was always considered to be the visual stimulus. Parameters a and b were found by minimizing the mean-squared error (Press *et al.*, 1992). We took the localisation error as the value of the residuals between data and fit.

Statistics

Statistical significance of a difference between two two-dimensonal (reaction time vs. localisation error) distributions was assessed by a two-dimensional Kolmogorov–Smirnov test. The means of two reaction time distributions were compared with a *t*-test, while the variances of localisation errors were compared with an *F*-test. We took $P = 0.05$ as the accepted level of significance.

Results

AV integration of aligned stimuli in aligned experiment

Synchronous presentation of spatially aligned audiovisual stimuli in the vertical plane led to faster and more accurate AV-evoked head saccades in the AV-100/0 experiment (only AV-aligned stimuli) than did presentation of unisensory stimuli. Figure 2A shows a representative example of the AV integration properties in this AV-100/0 experiment: the two-dimensional distributions (reaction time vs. localisation error) of A (blue), V (red) and aligned AV responses (green) of subject RM are compared to each other. The A responses were faster than the V responses, as the A distribution is systematically shifted to the left of the V distribution, but the V responses were clearly more accurate. However, the AV responses were, on average, the fastest and most accurate ($K\text{-}S_{791,791} = 0.59$, $P = 5.89 \times 10^{-83}$; $K\text{-}S_{761,792} = 0.26$, $P = 1.76 \times 10^{-16}$). This improvement was generally observed: five out of seven subjects had shorter AV reaction times than both A and V reaction times (Fig. 2B; *t*-test, $P \ll 0.001$; exceptions, with $P > 0.05$: subject MA for A; subjects MM and PB for V). The variance of the AV localisation error (Fig. 2C) was always lower (thus higher precision) than the variance in the A error (*F*-test, $P \ll 0.001$), while the AV error variance did not differ from the V error variance (*F*-test, $P > 0.05$), with the exception of subject PB ($F_{796,795} = 0.74$, $P = 2.2 \times 10^{-5}$).

We have termed this type of multisensory integration the ‘best of both worlds’ effect (Corneil *et al.*, 2002), as AV responses appeared to be as fast as, or faster than, A responses, but at visual localisation accuracy (equal mean) and precision (equal variance). Having established this typical improvement in performance for the basic spatially aligned AV stimuli when stimulus locations are confined to the vertical plane and responses are measured with head movements, we next quantified the effect of spatial disparity on AV integration.

Breakdown of AV integration by spatial disparity

A large spatial disparity between the A and V stimuli degraded the AV integration effect (Fig. 3) in the two experiments that contained auditory distractors: the AV-50/50 (50% distractors at > 45° disparity; Fig. 1E) and the AV-10/90 experiment (90% distractors at > 15° disparity; Fig. 1F). For small spatial disparities (= 20°, for which the V target and A distractor were in the same lower or upper hemifield), AV reaction times (Fig. 3A) and localisation errors (Fig. 3B) were not affected (*t*-test, $P > 0.05$ for all subjects). When the auditory distractor was presented in the hemifield opposite to the V target the size of the spatial disparity systematically delayed the AV responses, by up to ~38 ms for a disparity of 70° (Fig. 3A). A similar pattern was observed for the localisation errors (Fig. 3B): the error variance of the first saccade with respect to the visual target increased as AV disparity increased, with SD in localisation errors increasing up to ~16° for the largest spatial disparity.

These data are in good agreement with previous studies that reported a breakdown of AV integration when spatial disparity between the A and V stimulus increased (Harrington & Peck, 1998; Wallace *et al.*, 2004; Van Wanrooij *et al.*, 2009).

Effect of congruence-disparity distribution

So far, we have only shown the effects of the current stimulus properties on AV integration. We now address the question whether the ongoing distribution of AV disparities affects AV integration. This was investigated by presenting spatially aligned stimuli, either among other spatially aligned stimuli (AV-100/0 experiment), or among spatially disparate stimuli: AV-10/90, 10% aligned stimuli (Fig. 1F) and AV-50/50 experiment, 50% aligned stimuli (Fig. 1E). We found that spatially aligned AV stimuli elicited faster responses in the AV-100/0 experiment than in the AV-10/90 or AV-50/50 experiment, as exemplified for subject RM in Fig. 4A. The two-dimensional response distribution of localisation error vs. reaction time was faster and more precise in the AV-100/0 experiment (grey) than in the AV-10/90 experiment (black; $K\text{-}S_{789,72} = 0.32$, $P = 6.1 \times 10^{-5}$). The effect on reaction time is corroborated by the systematically lower value, by ~8 ms, in the AV-100/0 experiment than in the AV-10/90 and 50/50 experiments for all subjects (Fig. 4B; for nine out of 14 subjects and experiments $P < < 0.001$, while subject AK reacted faster in the AV-50/50 experiment: $t_{316} = 2.33$, $P = 0.02$). There was no significant improvement or decrement in localisation precision in the AV-100/0 experiment in any subject (*t*-test, $P > 0.05$).

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

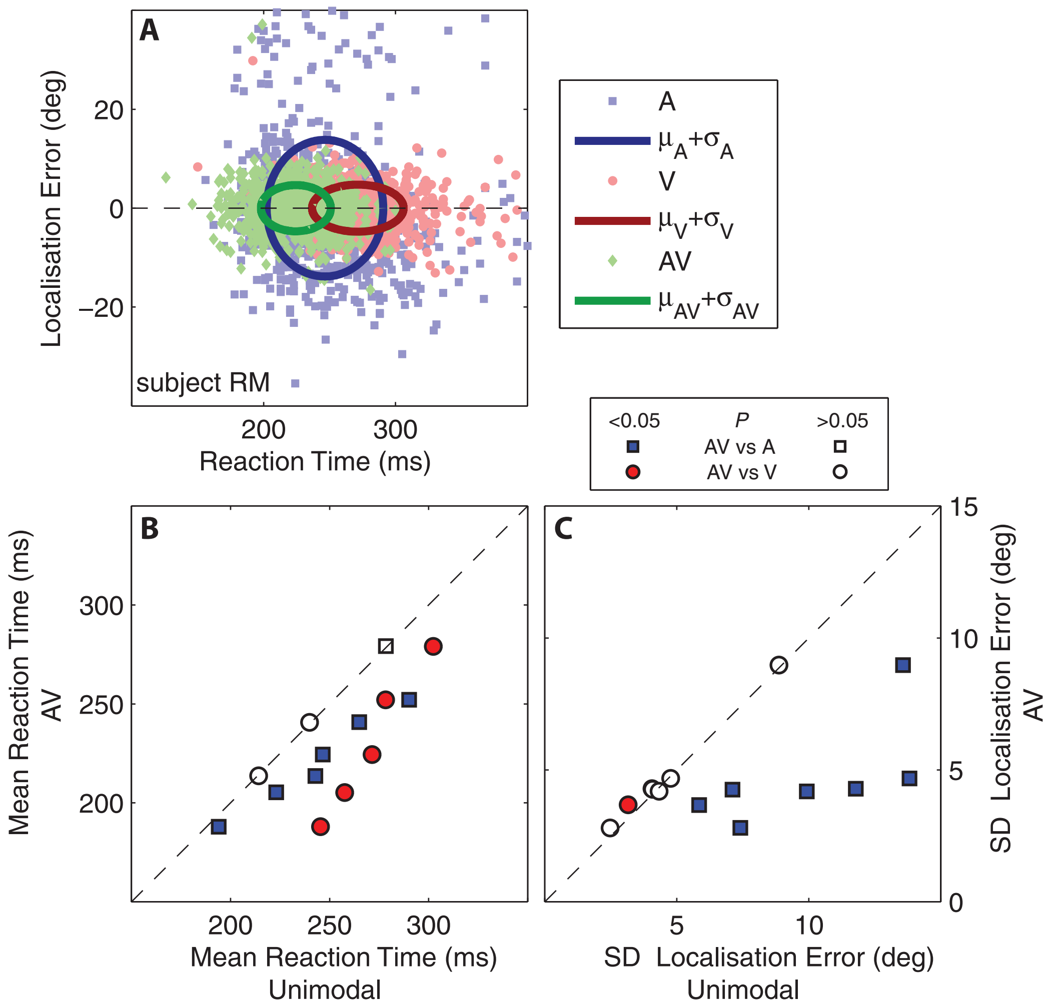

FIG. 2. AV integration in AV-100/0 experiment. Responses to unisensory (A, blue; V, red) and AV-aligned (green) responses. (A) Signed localisation error vs. reaction time for subject RM. Ellipses denote 1 SD around the mean. (B) Mean reaction time of AV-aligned responses vs. A and V unisensory responses for all subjects. (C) SD of localisation error for AV-aligned responses vs. A and V unisensory responses for all subjects. Open symbols, nonsignificant difference between unimodal and AV distributions ($P > 0.05$, t-test for reaction times, F-test for signed localisation errors).

Taken together these results imply that, when subjects can expect a disparate trial, an increase in reaction times is observed in the absence of an observable change in localisation errors.

Influence of previous trial disparity on reaction time

If the prior expectation of subjects is continuously updated, it is expected that the spatial configuration of a previous trial will influence the response to a novel audiovisual event, even if the novel stimuli themselves were presented at different locations. Because in both the AV-50/50 and AV-10/90 experiments trials were randomly interleaved, we had a large number of congruent and disparate trials that were preceded by either a congruent or a disparate trial. Current-trial spatial alignment had a large effect on AV integration (Fig. 3), but so did the audiovisual disparity of the previous trial.

Figure 5 shows the effect of the immediate trial history on the head-saccade reaction times when a currently congruent stimulus was preceded by either another congruent trial, or by a disparate trial (cf. Fig. 1D). Figure 5A shows the two possible current-trial configurations: either congruent (XC), or disparate (XD), regardless of the spatial alignment of previous trials (indicated by X). Clearly, the average reaction time in the current trial increased for a disparate target configuration (see also Fig. 3). In Fig. 5B we show the average reaction times when the current trial was congruent, whereas the previous trial could be either congruent (XCC), or disparate (XDC). Interestingly, the double-congruent condition yielded significantly faster reaction times (~8 ms) than trials in which the previous trial was disparate ($t_{17} = -4$, $P = 0.00094$). Note that this effect disappeared entirely when the current trial was disparate (Fig. 5C). In that case all reaction times were elevated, regardless of the alignment of previous trials (congruent, XCD; disparate, XDD). No effects were observed for the localisation errors (not shown). Hence, the target history concerning spatial alignment vs. spatial disparity influenced the reaction time of a current congruent trial in a highly nontrivial way. In the Discussion we present a simple theoretical account for this finding.

Discussion

The present study tested whether human subjects adaptively account for the expected alignment of audiovisual stimuli when programming a rapid head-orienting response. In conditions in which audiovisual

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

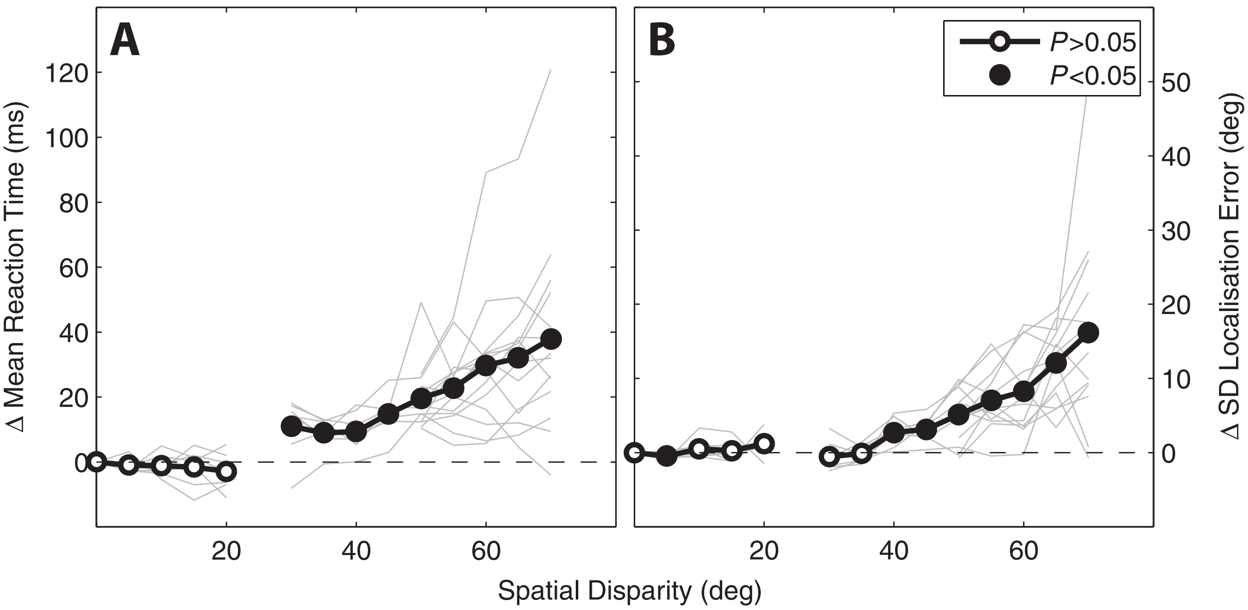

FIG. 3. Spatial AV disparity influences AV integration in the median plane. (A) Average reaction time as a function of spatial disparity. (B) Average localisation error SD as a function of spatial disparity. Reaction time and localisation error of aligned responses (spatial disparity 0°) was taken as baseline. Grey lines, different subjects and experiments; black bold lines, average over subjects and experiments; closed circles, significantly (t-test across subjects, $P < 0.05$) different from 0.

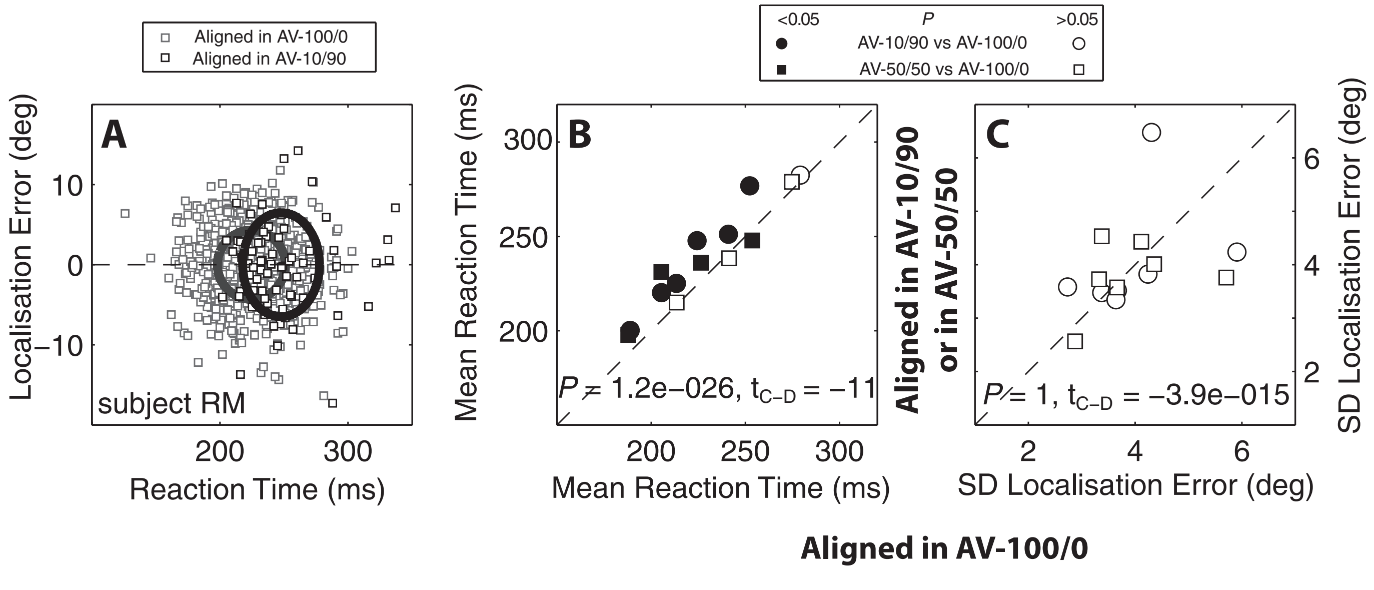

FIG. 4. Effect of disparity distribution on aligned AV responses. (A) Localisation error vs. reaction time of responses to aligned AV stimuli in the aligned AV-100/0 (grey) and AV-10/90 distractor (black) experiments for subject RM. Ellipses correspond to 1 SD around the mean. Note the shift toward shorter reaction times, and reduction in the variance in error when all stimuli in an experiment were aligned. (B) Average reaction time of AV-aligned responses in the AV-10/90 (circles) and AV-50/50 (squares) experiment vs. AV-aligned responses in the AV-100/0 experiment for all subjects. (C) SD in localisation error of AV-aligned responses in AV-10/90 and AV-50/50 experiments vs. AV-aligned responses in AV-100/0 experiment for all subjects. Closed symbols in panels B and C represent significant differences for a single subject; $P < 0.05$.

stimuli were spatially aligned within the median plane, reaction times were lower, accuracy was higher and endpoint variability was lower than in the unisensory evoked response statistics (Fig. 2). When stimuli were spatially disparate, the benefit of audiovisual integration broke down (Fig. 3). We measured the reaction time and accuracy of head saccades to simultaneously presented audiovisual stimuli, and explored whether there was a difference when the spatial stimulus statistics in the experiment changed. We hypothesized that, if the brain keeps track of the audiovisual congruency of prior trials to update its expectation of current stimulus alignment, we should observe an effect of stimulus history on the orienting responses.

Indeed, our main finding was that head saccades were systematically altered by the stimulus statistics: this was demonstrated by lower average reaction times for experiments with only spatially-aligned stimuli than for experiments in which the probability for spatial congruency was reduced (Fig. 4). Interestingly, sequences of trials that contained a larger proportion of congruent stimuli had lower reaction times than did disparate stimulus sequences (Fig. 5). The change in

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

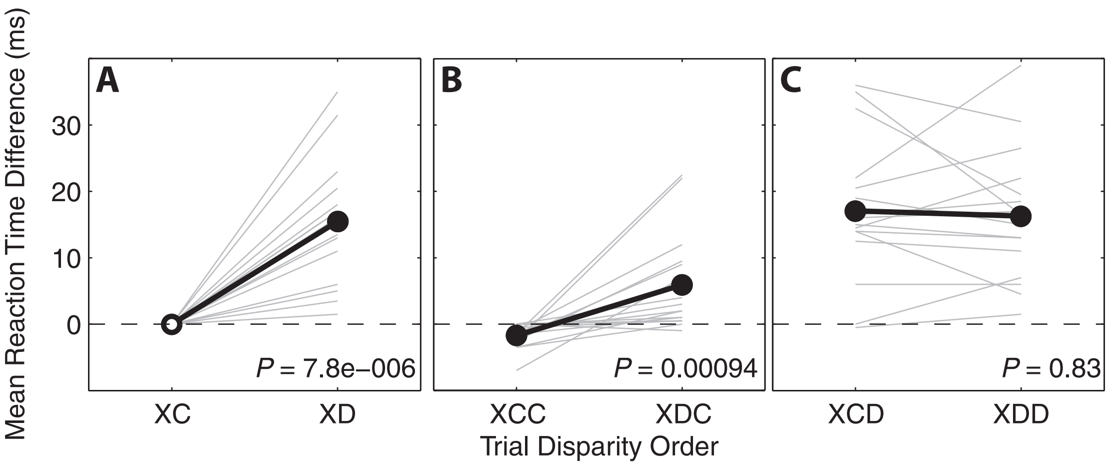

FIG. 5. Influence of trial history. (A) Differences in average reaction time for the current disparate trial (XD) vs. the baseline congruent trial (XC) regardless of the alignment of the previous trial (X). (B) Reaction times are compared for congruent trials that were preceded by another congruent trial (XCC), or by a disparate trial (XDC). The double-congruent condition yielded the fastest reaction times, even significantly faster than the general XC trials. (C) The improvement was not obtained when the current trial was disparate (XCD and XDD trial sequences). Grey lines, individual subjects and experiments; black bold line, average across subjects and experiments; closed circles, significant difference between the two sequences in the subplots ($P < 0.05$, *t*-test).

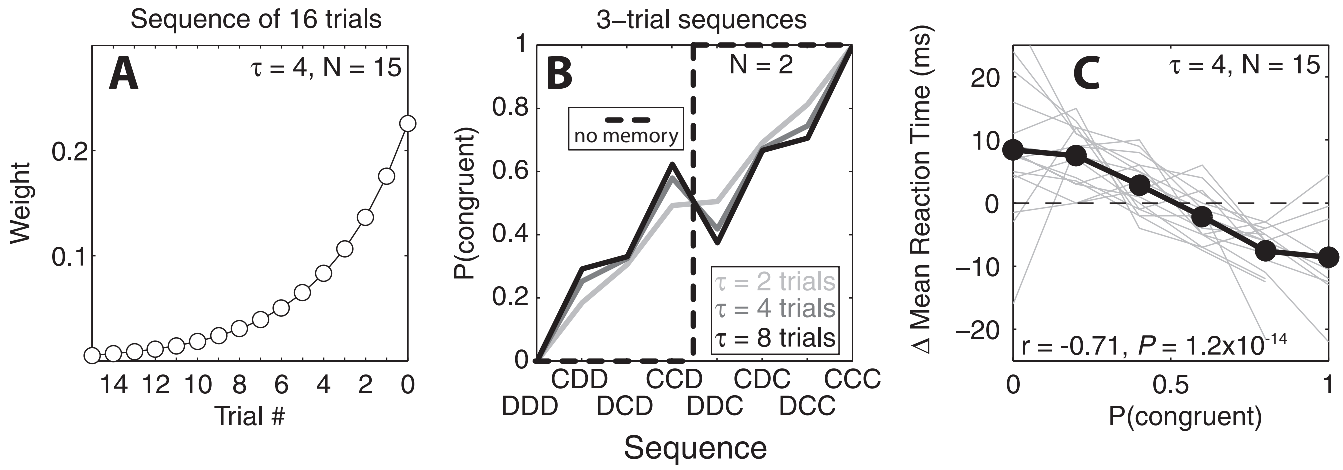

FIG. 6. Explanatory model for dynamic probability estimation of audiovisual congruency. (A) Memory weights for a sequence of 16 trials follow an exponential decay with time constant, τ = 4 trials. (B) Example *P(congruent)*, for 3-trial sequences for time constants of τ = 2, 4 and 8 trials. Note that if multisensory integration would have no memory of previous trials, the prediction would follow a step function (dashed line), with *P(congruent)* = 1 for all XXC trials and *P(congruent)* = 0 for all XXD trials. (C) Average reaction time plotted against *P(congruent)* for N = 15, τ = 4, for all subjects and experiments. Reaction times and *P(congruent)* of individual trials have been binned on the basis of 0.2-wide windows of *P(congruent)*. Note that reaction time is clearly modulated by *P(congruent)*. Grey lines – individual subjects and experiments; black bold line – average across subjects and experiments.

reaction time depended systematically on the particular order of congruent and disparate stimuli of previous trials, regardless of their actual spatial locations. This suggests that the brain indeed constructs a dynamic expectation regarding the spatial alignment of novel stimuli, which readily adjusts multisensory integration on the basis of recently acquired evidence.

Audiovisual integration

In line with earlier research (Hughes *et al.*, 1994; Nozawa *et al.*, 1994; Frens *et al.*, 1995; Goldring *et al.*, 1996; Colonius & Arndt, 2001; Corneil *et al.*, 2002; Van Wanrooij *et al.*, 2009), we found that simultaneous presentation of spatially aligned stimuli in the median plane also systematically reduced the reaction times and accuracy for head saccades. Thus, the 'best-of-both-worlds' principle (Corneil *et al.*, 2002; Van Wanrooij *et al.*, 2009) also applies to head movements, which are much slower and more variable than gaze saccades, especially in the vertical plane (Goossens & Van Opstal, 1997). This is a strong indication that the effects of audiovisual integration are real, and independent of the particular pointer used, of the stimulus environment (few or many targets, with or without a perturbing noisy background) and of the response dimension (one-dimensional, two-dimensional, horizontal or vertical).

Interestingly, our data indicate that the sensorimotor system seems to interpret physical AV disparities of up to 20–30° as being 'congruent' (Fig. 3). This spatial resolution is similar to the value reported for target averaging in the visuomotor system (Ottes *et al.*, 1984, 1985). In their experiments visual double-stimuli with spatial disparities of up to 30° in direction typically resulted in averaged saccade responses. Recently, we also showed similar averaging effects

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

for head saccades to auditory double stimuli presented in the median plane (Bremen *et al.*, 2009). Our finding strongly suggests a general phenomenon in which two stimuli are fused based on their perceived proximity in order to generate a rapid saccadic response, irrespective of the modality of those stimuli.

Effect of stimulus statistics on saccade generation

Note that we presented stimuli at randomly selected locations within the median plane over a large range (70°). In other words, it was highly unlikely that a particular spatial stimulus configuration would be repeated in the next trial. The only attribute that was systematically varied in the experiments was the probability for spatial congruency of the stimuli. Apparently, the mechanisms that subserve multisensory integration are able to extract and use this particular stimulus statistic.

The effect of stimulus statistics on saccade reaction times to visual stimuli has also been studied by Dorris *et al.* (2000) during single-unit recording of saccade-related activity in the monkey superior colliculus. In that study, the target could only occupy one of two possible locations (left vs. right) but the saccade reaction times, as well as the superior colliculus premotor activity levels, systematically varied with the trial history in much the same way as observed in the current study. For example, reactions times were shortest (and premotor activity was highest) after a sequence of identical target locations, while reaction time increased (activity decreased) when the sequence alternated erratically between left and right.

Their data showed that the history effect occurred for the two particular stimulus locations employed, but they did not test whether the effect would transfer to other locations. Our study indicates that the reaction time effect on audiovisual integration generalises across all locations when the subject is tested for a large range of targets. Hence, the gaze control system is able to extract the relevant parameter from the trial statistics, which in our case was audiovisual spatial disparity. In line with the results of Dorris *et al.* (2000), and based on multisensory integration studies in anesthetised preparations (Meredith & Stein, 1986a,b; Meredith *et al.*, 1987), we conjecture that AV integration would be reflected in the activity of saccade-related cells of the superior colliculus. It would therefore be interesting to verify whether and how the rules of multisensory integration that emerge from our study would be reflected in the sensory and/or preparatory activity epochs of these neurons.

Stimulus statistics and adaptive coding

In our experiments, we changed a particular stimulus statistic (i.e. the probability of audiovisual congruency) and observed a change in the ensuing behavioural responses (reaction times). Because the AV stimuli were presented in randomised order, and our subjects reacted as fast as possible, we believe that it is highly unlikely that subjects somehow adopted a conscious strategy by which they could purposefully influence their reaction time statistics. We therefore conjecture that the observed phenomenon is due to an automatic, bottom-up (exogenous) neural process rather than to cognitive, top-down (endogenous) factors. Interestingly, dynamic adjustments of neural responses according to stimulus statistics have also been reported for the early auditory pathway of anesthetised guinea pig inferior colliculus (Dean *et al.*, 2005), and even for the cat auditory nerve (Wen *et al.*, 2009), as well as for the fly visual system (Brenner *et al.*, 2000) and the rat barrel cortex (Maravall *et al.*, 2007). These studies all showed that adaptive spike-rate changes to variations in the stimulus statistics can be rapid and are driven by automatic stimulus-driven (exogenuous) processes. We are unaware of any neurophysiological study describing spike-rate adaptation to multisensory stimulus statistics, and it would therefore be interesting to search for potential neural mechanisms of the behavioural changes reported in this study.

Model for reaction time modulation by prior likelihood estimation

The findings shown in Figs 4 and 5 strongly suggest that audiovisual integration has a dynamic component that depends on the evidence for stimulus congruency as acquired from prior experience. To explain these results within a probabilistic framework we constructed a simple model that estimates the likelihood that audiovisual events may be congruent by weighting prior evidence. We further assume that the likelihood, P(congruent), modulates the saccade reaction time: the higher the likelihood, the earlier the response onset. In the model the prior probability of a congruent stimulus configuration is found by continuously updating the weighted averaged probability for stimulus alignment. In this model the occurrence of a disparate stimulus configuration is given weight zero (i.e. a zero *post hoc* probability of being aligned). For simplicity, we took the dynamic probability for congruent stimulus configurations to follow an exponential decay (with a time constant τ, and a memory n = 15 trials):

$$P(\text{congruent}; n=0) = \sum_{n=0}^{N} w(n) \cdot P(\text{congruent}|\text{congruent}, \text{incongruent})$$
$$\text{with } w(n) = w_0 \cdot \exp(-n/\tau)$$
$$P(\text{congruent}|\text{congruent}) = 1 \text{ and } P(\text{congruent}|\text{incongruent}) = 0$$

where w_0 is a normalization factor, which is 0.2253 for τ = 4 and n = 15. Such a model favours the most recent trials, and mimics a leaky memory (Fig. 6A). For example, the contribution of an aligned trial that occurred six trials before the current trial is given a weight $w(6)$ = 0.0503, for τ = 4 and n = 15. In Fig. 6B we show the probabilities when the memory extends over n = 2 trials, for all eight possible triplet sequences, for different decay time constants, τ = 2, 4 and 8 trials, respectively. Note that the model predicts that the trial sequence CCD (current trial disparate) has a higher probability for alignment (shorter reaction time) than DDC (current trial congruent) for time constants of 4 and 8. Applying the model with τ = 4 and n = 15 to our data, we show a good correspondence between the average change in reaction times (positive, increase; negative, decrease) and the estimated probability for congruent stimuli for the great majority of subjects (binned in 0.2-wide windows; Fig. 6C). A time constant of 4 trials yielded the largest r^2-value (r_{87} =−0.71, $P = 1.2 \times 10^{-14}$).

In summary, our experiments provide support evidence that the brain makes a dynamic evaluation of the multisensory scene to program a rapid orienting response that may or may not be based on multisensory integration. Such a strategy is particularly useful in unpredictable and complex environments where the statistics of auditory and visual stimuli may continuously vary.

Acknowledgements

We thank Sharon Smits for help in the earlier phase of this study. We are grateful to Hans Kleijnen, Dick Heeren and Stijn Martens for valuable technical assistance. This research was supported by a Marie Curie Early Stage Training Fellowship of the European Community's Sixth Framework Program (MEST-CT-2004-007825; P.B.), a VICI grant within the Earth and Life Sciences of NWO (ALW 865.05.003; A.J.V.O., M.M.V.W.) and the Radboud University Nijmegen (A.J.V.O.)

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Abbreviations

A, auditory; AV, audiovisual; AV-10/90, audiovisual 10%-aligned/90%-distractor experiment; AV-100/0, audiovisual 100%-aligned/0%-distractor experiment; AV-50/50, audiovisual 50%-aligned/50%-distractor experiment; C, congruent; D, disparate; LED, light-emitting diode; V, visual.

References

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

Anastasio, T.J., Patton, P.E. & Belkacem-Boussaid, K. (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput.*, **12**, 1165–1187.

Bell, A.H., Meredith, M.A., Van Opstal, A.J. & Munoz, D.P. (2005) Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *J. Neurophysiol.*, **93**, 3659–3673.

Bremen, P., Van Wanrooij, M.M. & Van Opstal, A.J. (2009) Pinna cues determine orienting response modes to synchronous sounds in elevation. *J. Neurosci.*, in press. **130**, 194–204.

Brenner, N., Bialek, W. & Ruyter van Steveninck, R. (2000) Adaptive rescaling maximizes information transmission. *Neuron*, **26**, 695–702.

Colonius, H. & Arndt, P. (2001) A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys*., **63**, 126–147.

Corneil, B.D., Van Wanrooij, M., Munoz, D.P. & Van Opstal, A.J. (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J. Neurophysiol.*, **88**, 438–454.

Dean, I., Harper, N.S. & McAlpine, D. (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.*, **8**, 1684–1689.

Dorris, M.C., Pare, M. & Munoz, D.P. (2000) Immediate neural plasticity shapes motor performance. *J. Neurosci.*, **20**, RC52.

Ernst, M.O. (2005) A Bayesian view on multimodal cue integration. In Knoblich, G., Grosjean, M., Thornton, I. & Shiffrar, M. (Eds), *Human Body Perception from the Inside Out*. Oxford University Press, New York, pp. 105–131.

Frens, M.A. & Van Opstal, A.J. (1998) Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res. Bull.*, **46**, 211–224.

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

Goldring, J.E., Dorris, M.C., Corneil, B.D., Ballantyne, P.A. & Munoz, D.P. (1996) Combined eye-head gaze shifts to visual and auditory targets in humans. *Exp. Brain Res.*, **111**, 68–78.

Goossens, H.H. & Van Opstal, A.J. (1997) Human eye-head coordination in two dimensions under different sensorimotor conditions. *Exp. Brain Res.*, **114**, 542–560.

Harrington, L.K. & Peck, C.K. (1998) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp. Brain Res.*, **122**, 247–252.

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

Hughes, H.C., Reuter-Lorenz, P.A., Nozawa, G. & Fendrich, R. (1994) Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol*., **20**, 131–153.

Hughes, H.C., Nelson, M.D. & Aronchick, D.M. (1998) Spatial characteristics of visual-auditory summation in human saccades. *Vision Res.*, **38**, 3955–3963.

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

Maravall, M., Petersen, R.S., Fairhall, A.L., Arabzadeh, E. & Diamond, M.E. (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS Biol.*, **5**, e19 323–e19 334.

Meredith, M.A. & Stein, B.E. (1986a) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.*, **365**, 350–354.

Meredith, M.A. & Stein, B.E. (1986b) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.*, **56**, 640–662.

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

Nozawa, G., Reuter-Lorenz, P.A. & Hughes, H.C. (1994) Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades. *Biol. Cybern.*, **72**, 19–34.

Ottes, F.P., Van Gisbergen, J.A.M. & Eggermont, J.J. (1984) Metrics of saccade responses to visual double stimuli: two different modes. *Vision Res.*, **24**, 1169–1179.

Ottes, F.P., Van Gisbergen, J.A.M. & Eggermont, J.J. (1985) Latency dependence of colour-based target vs nontarget discrimination by the saccadic system. *Vision Res.*, **25**, 849–862.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vettering, W.T. (1992) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge MA, USA.

Robinson, D.A. (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.*, **10**, 137–145.

Sato, Y., Toyoizumi, T. & Aihara, K. (2007) Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.*, **19**, 3335–3355.

Stein, B.E. & Meredith, M.A. (1993) *The Merging of the Senses*. MIT, Cambridge, MA.

Van Wanrooij, M.M., Bell, A.H., Munoz, D.P. & Van Opstal, A.J. (2009) The effect of spatial–temporal audiovisual disparities on saccades in a complex scene. *Exp. Brain Res.*, **198**, 425–437.

Wallace, M.T., Meredith, M.A. & Stein, B.E. (1998) Multisensory integration in the superior colliculus of the alert cat. *J. Neurophysiol.*, **80**, 1006–1010.

Wallace, M.T., Roberson, G.E., Hairston, W.D., Stein, B.E., Vaughan, J.W. & Schirillo, J.A. (2004) Unifying multisensory signals across time and space. *Exp. Brain Res.*, **158**, 252–258.

Wen, B., Wang, G.I., Dean, I. & Delgutte, B. (2009) Dynamic range adaptation to sound level statistics in the auditory nerve. *J. Neurosci.*, **29**, 13797–13808.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd