The effects of stereo disparity on the behavioural and electrophysiological correlates of perception of audio-visual motion in depth

Neil R. Harrison[1], Sian Witheridge [2], Alexis Makin[2], Sophie M. Wuerger[2], Alan J. Pegna[3], Georg F. Meyer[2]*

Affiliations

[1] Liverpool Hope University, Dept of Psychology, Hope Park, Liverpool, L16 9JD, UK

[2] University of Liverpool, Dept of Psychological Sciences, Eleanor Rathbone Building, L69 7ZA, UK

[3] Laboratory of Experimental Psychology, Faculty of Psychology and Educational Science, University of Geneva, and Neurology Clinic, Geneva University Hospital, Geneva, Switzerland

* corresponding author (georg@liv.ac.uk, tel. +44 151 7942579)

Abstract

Motion is represented by low-level signals, such as size-expansion in vision or loudness changes in the auditory modality. The visual and auditory signals from the same object or event may be integrated and facilitate detection. We explored behavioural and electrophysiological correlates of congruent and incongruent audio-visual depth motion in conditions where auditory level changes, visual expansion, and visual disparity cues were manipulated. In Experiment 1 participants discriminated auditory motion direction whilst viewing looming or receding, 2D or 3D, visual stimuli. Responses were faster and more accurate for congruent than for incongruent audio-visual cues, and the congruency effect (i.e., difference between incongruent and congruent conditions) was larger for visual 3D cues compared to 2D cues. In Experiment 2, event-related potentials (ERPs) were collected during presentation of the 2D and 3D, looming and receding, audio-visual stimuli, while participants

detected an infrequent deviant sound. Our main finding was that audio-visual congruity was affected by retinal disparity at an early processing stage (135 – 160 ms) over occipito-parietal scalp. Topographic analyses suggested that similar brain networks were activated for the 2D and 3D congruity effects, but that cortical responses were stronger in the 3D condition. Differences between congruent and incongruent conditions were observed between 140 – 200 ms, 220 – 280 ms, and 350 – 500 ms after stimulus onset.

Introduction

Approaching (or 'looming') objects, which often necessitate immediate action to avoid collision or escape predation, can be represented in both the visual and auditory modalities. Vision and hearing use different mechanisms to detect motion in the depth plane; vision relies on cues such as retinal expansion and binocular disparity, whereas the auditory system utilises intensity changes (Bach, Neuhoff, Perrig, & Seifritz, 2009; Regan & Gray, 2000). It is well known that the integration of sensory cues is governed by several basic principles, including the requirement for spatial and temporal congruence (Stein & Meredith, 1993; Meyer et al., 2005). These basic principles are also evident at a behavioural level; for example (Meyer and Wuerger, 2001). Cappe, Thut, Romei, and Murray (2009) showed that reaction times to looming objects are reduced by combining information from the visual and auditory modalities.

For unimodal visual and auditory stimuli, there is a specific processing bias that causes looming cues to be perceived as more salient than receding cues (e.g. Bach et al., 2009; Franconeri & Simons, 2003). This bias has been explained in evolutionary terms, with clear adaptive advantages for the processing of looming stimuli associated with collision avoidance and escape from predation (Franconeri & Simons, 2003). The processing bias for looming

signals has also been demonstrated for multisensory cues, where multisensory motion detection was faster and more accurate for congruent, multisensory looming cues, compared to receding, incongruent or unimodal signals (Cappe, Thut, Romei, & Murray, 2009). Moreover, Harrison (2012) reported that the looming bias extended to the phenomenon of dynamic visual capture, in that the visual 'capture' of auditory motion direction was stronger for looming than receding stimuli, as measured by the accuracy of motion direction discrimination for tones.

Typically, experiments investigating motion in the depth plane have used linearly expanding geometric objects to induce the perception of visual motion in depth (e.g. Cappe, Thut, Romei & Murray, 2009; Harrison, 2012). It is an open question whether binocular disparity cues also convey reliable information about motion in depth. González, Allison, Ono, and Vinnikov (2010) argue that changes in relative disparity and vergence, elicited by changing disparity, are effective cues to motion in depth, while Erkelens and Collewijn (1985) and Regan, Erkelins, and Collewijn (1986) argue that vergence changes alone do not induce a sensation of motion in depth. Ogawa and Macaluso (2013) found that the addition of stereo (3D) depth cues did not influence discrimination of audio motion consistent with arguments for the limited importance of binocular disparity in collision avoidance (Regan & Gray, 2000). Despite this, in the same study (Ogawa & Macaluso, 2013), fMRI revealed enhanced activation in region V3 as well as increased connectivity to auditory cortex for audio-congruent, stereo-looming motion, consistent with depth cues being used in audio-visual motion integration. This is consistent with recent neuroimaging data providing evidence for the integration of binocular disparity and relative motion cues in the dorsal visual area V3B (Ban, Preston, Meeson, & Welchman, 2012).

In light of these inconsistencies, the impact of disparity on dynamic visual capture merits further investigation, in particular in relation to the timing of neural activation in response to

audio-visual 3D cues. Event-related potentials (ERPs) are ideal measures to assess the timing of neural processes, given their temporal resolution on the scale of milliseconds, and the latencies of neural correlates of multisensory mechanisms have been investigated in numerous ERP studies.

In the current ERP study, we compare bimodal signals that differ in terms of their directional congruity: that is signals that both travel in the same direction (congruent condition), or in opposite directions (incongruent condition) to investigate the specific neural mechanisms involved in the multisensory processing of dynamic depth cues.

Enhanced negative EEG amplitudes have been reported for congruent AV stimuli occurring at latencies of around 250ms over temporal and fronto-central regions (Bonath et al., 2007; Busse, Roberts, Crist, & Weissman, 2005; Proctor & Meyer, 2011). Selective responses for incongruent audio-visual stimuli have also been reported at latencies beyond 300 ms (e.g., Zimmer, Itthipanyanan, Grent-'t-Jong, & Woldorff, 2010; Proctor & Meyer, 2011), and may reflect an N400-like effect (Kutas & Hillyard, 1989; Diaconescu, Alain, & McIntosh, 2011; Szűcs & Soltész, 2007).

The above studies have generally used stationary cues; less is known about the timing of neural processing involved in the multisensory perception of dynamic cues, and in particular about the timing of neural processes related to the processing of audio-visual motion in depth. Cappe, Thelen, Romei, Thut, & Murray (2012) investigated the timing of neural responses to audio-visual depth motion cues using ERPs, and found interactions starting around 75 ms for audio-visual looming conditions. However, Cappe et al. (2012) used only two-dimensional cues and so did not assess the effect of stereo disparity in their study.

The current research aims to investigate behavioural responses (Experiment 1) and the time-course of neural activity (Experiment 2) related to the perception of audio-visual motion in

depth, and in particular whether the perception of audio-visual motion in depth is affected by stereo disparity of the visual signals. Both experiments compare congruent (i.e., both signals moving in the same direction) and incongruent (i.e., signals moving in opposite directions) audio-visual cues.

In Experiment 1 participants discriminated looming from receding apparent auditory motion signals while simultaneously observing task-irrelevant, directionally congruent and incongruent, looming and receding, 2D (visual expansion) and 3D (visual expansion and disparity/vergence cues) visual stimuli. We predicted that auditory motion discrimination would be faster and more accurate for congruent rather than incongruent AV conditions and that this effect would be mediated by motion depth direction and the presence of disparity cues. Specifically, it was expected that responses associated with looming visual cues would be faster and more accurate than receding. For disparity, assuming a stronger signal as evidenced by enhanced activation for congruent 3D cues (Ogawa & Macaluso, 2013), it was expected that increased speed and accuracy would be associated with 3D relative to expanding 2D visual signals.

Experiment 2 aimed to explore the timing of crossmodal neural mechanisms related to the perception of audio-visual motion in depth, and in particular how stereo disparity modulated the ERP responses. An oddball task was employed where participants detected an infrequent deviant sound. Only ERPs to the frequent signals were analysed, so that the electrophysiological results would not be confounded by motor-related neural activity. We expected to observe early (i.e., < 200 ms) effects of directional congruence and stereo disparity over posterior scalp, and at later stages of processing, enhanced negativity for incongruent compared to congruent stimuli reflecting conflict between opposing motion directions.

Experiment 1: Behavioural Responses

Methods

Participants

Thirteen participants took part in this experiment (4 male; mean age = 26.9 years; SD = 7.3). All were staff or students at the University of Liverpool. All participants reported normal or corrected-to-normal vision and normal hearing. All had a stereo acuity threshold better than 100 arcsec, with mean acuity at 45.38 (SD = 11.26) arcsec; comparable to population normal means (Kim, Yang, Kin, Lee & Hwang, 2011), as measured by a Stereo Fly test, SO-001 (Stereo-Optical, 2007).

Stimuli, materials and apparatus

Visual Stimuli

A selection of nine easily recognisable images of objects was taken from a database of 209 stimulus images, courtesy of Michael J. Tarr, http://www.tarrlab.org/. Images were similar in size, resolution (450 x 450 pixels) and orientation (30 degree x-axis rotation) and were inserted onto a blue-white textured background ('hockey-ice.jpg', from www.psdgraphics.com).
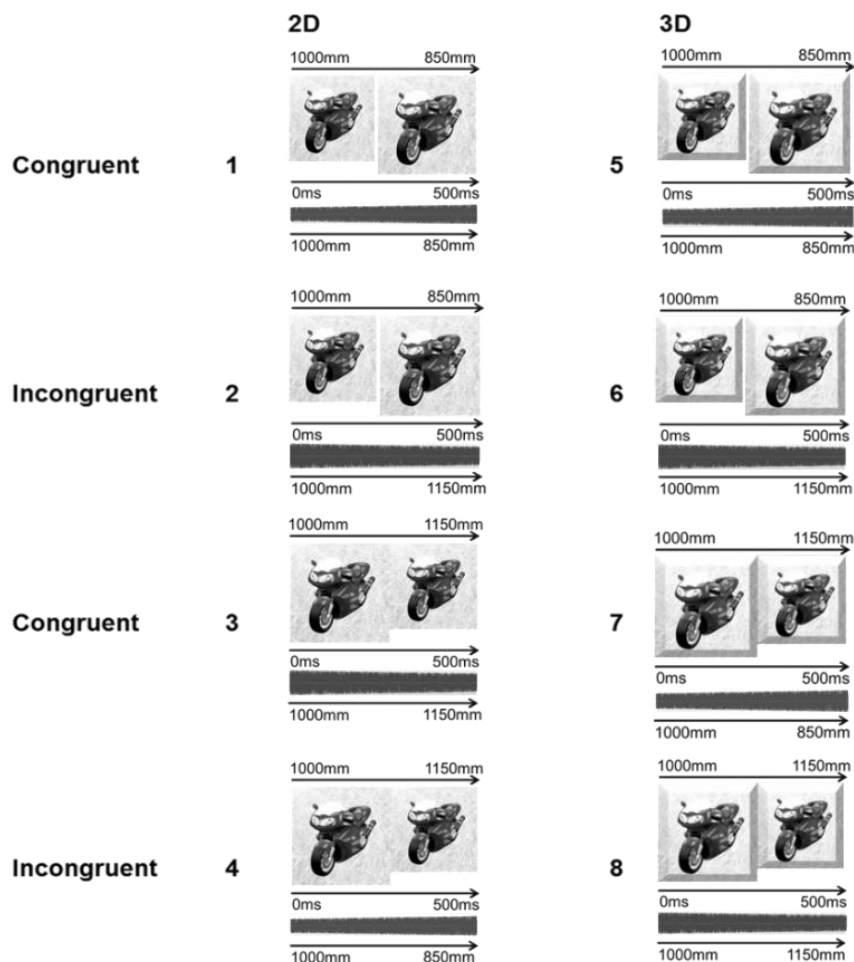
Fig. 1. *Schematic illustrating the eight conditions of both experiments. Congruence was manipulated by presenting directionally matched or mismatched audio-visual combinations. All size and intensity changes were at 15%, started from 100 cm viewing distance and changed over 500ms.*

Apparent depth motion was induced by changing the size of objects by 15%. Motion always started at 100 cm viewing distance (the plane of the monitor) and was linear over a duration of 500 ms. Stereo disparity was generated by interleaving horizontally alternative pixels at disparities corresponding to viewing distance changes of 15% from the start point (1000 mm) for 3D stereo. The background image had a constant disparity of 52mm (infinity). For the 2D condition the 'left and right image' was identical. At the signal onset visual images extended

over 20 deg (horizontal) and 11 deg (vertical) visual angle while the monitor extended over

28.5 deg (horizontal) and 15.8 deg (vertical) visual angle.

Auditory Stimuli

The amplitude of broadband (white) noise signals was modulated to correspond to the size

change of visual stimuli (15%). The size change used in all experiments was selected on the

basis of a preliminary threshold detection experiment: 23 participants completed a 2-forced

choice discrimination task for looming and receding broadband noise, corresponding to

distance changes of between 5% and 40%. A 15% intensity change reflected average

accuracy scores of approximately 75% for looming and 65% for receding signals.

Audio-visual stimuli

AV pairings, overlaid using video editing software, were quasi-randomly interleaved in a

block of 216 trials (27 per condition); all clips were 0.5s duration. A fixation frame of 3.5s

preceded trials, which began with an animation of 0.5s in which a blue cross of 110 x 90

pixels loomed or receded by 15%, then held stationary for 3s at 1000mm in order to

normalise vergence eye movements (Figure 2). The cross was blue on the same 'hockey-ice'

background as objects. A practice block was constructed containing twelve AV trials and

fixation animations.
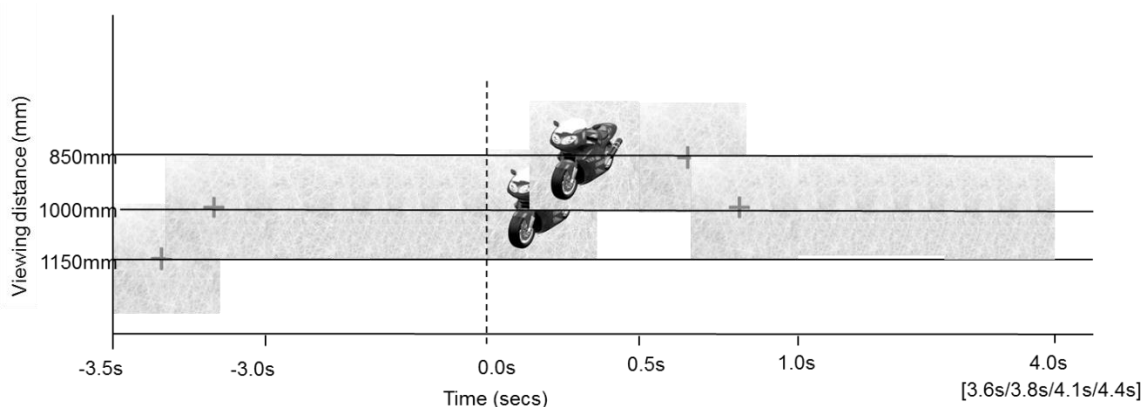
Fig. 2. *Schematic illustrating vergence normalisation for a visual receding to looming trial.*
Times in rectangular parenthesis represent jittered lengths used in Experiment 2.

Apparatus

Dichotic AV movies were presented in a darkened sound-proof booth, on a 23 inch passive
3D monitor (LG D2342P) at 1.0 m viewing distance. Auditory signals, peak amplitude
70dB(A) at the listener's position, were played through loudspeakers placed directly
underneath the monitor. Plastic polarising glasses were worn throughout.

Audio Training Task

An audio training task was completed prior to the main experiment to familiarise participants
to audio depth-motion discrimination. This involved a two-forced choice task, with feedback,
for discrimination of looming and receding sounds. The training was run in Psychopy
(Version 1.73; see Peirce, 2007), on a laptop computer, with sounds received via headset,
peaking at 70 dB(A).

Results

Reaction times and accuracy scores were analysed separately in 2 (trial-block) x 2 (visual
motion) x 2 (congruence) x 2 (disparity) repeated measures analyses of variance (ANOVAs).
Overall trial-block results will be described first to assess practice effects, followed by results
for response latencies and then accuracy. Responses across conditions in block two were
significantly faster (1084.21 [±115.47 *SD*] ms) than block one (1178.47 [±104.48 *SD*] ms;
$F(1, 12) = 4.961$, $p = 0.05$). There was no difference in accuracy between block one (67.45
[±3.50 *SE*]) and block two (66.99 [±3.30 *SE*]), ($F(1,12) = .02$, $p = .69$), see tables 1 and 2 for
detailed descriptive and inferential statistics.

Response latencies in congruent trials were significantly shorter (1055.47 [±*105*.83 *SE*] ms) than incongruent (1207.22 [±113.41 *SE*] ms) trials, (F(1, 12) = 16.18, p = 0.002). No significant interactions were observed, counter to predictions for response times relating to modulation of congruence by either motion or disparity cues.

Table 1. *Mean and standard error (SE) of motion discrimination accuracy (% correct) and reaction times (ms) for looming and receding, 2D and 3D, congruent and incongruent conditions.*

	Condition	Mean % correct (±*SE*)	Mean RT (ms) (±*SE*)
Congruent	Looming 2D (VLAL)	82.48 (2.36)	1037 (114.6)
	Looming 3D (VLAL)	82.69 (2.33)	1063 (104.8)
	Receding 2D (VRAR)	84.76 (2.78)	1064 (110.3)
	Receding 3D (VRAR)	80.62 (2.74)	1057 (101.9)
Incongruent	Visual Looming 2D (VLAR)	49.58 (4.99)	1168 (108.6)
	Visual Looming 3D (VLAR)	47.01 (5.23)	1214 (114.4)
	Visual Receding 2D (VRAL)	59.54 (5.52)	1202 (115.6)
	Visual Receding 3D (VRAL)	56.55 (5.57)	1245 (123.5)

Accuracy of motion discrimination was significantly higher for congruent (82.63 [±2.26 *SE*] %) relative to incongruent (53.17% [±5.07 *SE*] conditions ($F(1, 12) = 47.05$, $p < 0.001$). The effect of congruence on accuracy was modulated by motion direction ($F(1,12) = 4.56$, $p = 0.05$). Simple effects of congruence on accuracy were significant for both looming ($F(1,12) = 43.05$, $p < 0.0001$) and receding visual motion ($F(1,12) = 15.85$, $p = 0.002$). We subtracted accuracy rates in incongruent conditions from accuracy rates in congruent conditions to obtain a measure of the congruency effect. We found marginally significant enhanced accuracy for congruent (35.33 % [±5.38 *SE*]) over incongruent (20.87 % [±5.24 *SE*]) conditions in looming relative to receding motion conditions ($t(12) = 2.14$, $p = 0.054$).

The effect of congruence on accuracy was also found to be modulated by disparity of the visual stimuli ($F(1,12) = 13.37$, $p = 0.003$). Both 2D and 3D were associated with significant simple effects of congruence (2D: $F(1,12) = 40.19$, $p < 0.001$ and 3D: $F(1,12 = 50.73$, $p < 0.001$) but there was a significantly larger congruency effect (congruent minus incongruent) for 3D (30.91 % [±4.34 *SE*]) relative to 2D (25.28 % [±3.99 *SE*]) conditions ($t(12) = 3.66$, $p = 0.003$).

Table 2: summary inferential statistics for reaction time and accuracy ANOVA analyses

Value	Reaction Times		Accuracy	
	F(1,12)	Sig.	F(1,12)	Sig.
trial_block (1 vs 2)	**4.951**	**.046**	.164	.692
visual_motion (looming vs receding)	1.766	.209	1.835	.200
Congruence (congruent vs incongruent)	**16.186**	**.002**	**47.052**	**.000**

Disparity (2D vs 3D)	3.689	.079	.001	.972
trial_block * visual_motion	1.578	.233	.740	.407
trial_block * congruence	.038	.848	**6.787**	**.023**
visual_motion * congruence	.633	.442	**4.560**	**.054**
trial_block * visual_motion * congruence	.260	.619	.047	.831
trial_block * disparity	.231	.639	1.187	.297
visual_motion * disparity	.252	.625	.021	.886
trial_block * visual_motion * disparity	1.587	.232	**7.634**	**.017**
congruence * disparity	.230	.640	**13.369**	**.003**
trial_block * congruence * disparity	.615	.448	.475	.504
visual_motion * congruence * disparity	.392	.543	.073	.791
trial_block * visual_motion * congruence * disparity	.122	.733	3.217	.098

Experiment 2: Event-related potentials

Methods

Participants

Fourteen participants took part in this study (3 male, mean age 24 years 5 months ±*SD* 4 years, 4 months). Data from one participant was unusable due to equipment failure and two further participants were excluded due to excessive electrode impedances.

Stimuli, materials and apparatus

Visual stimuli

All images were identical to those used in Experiment 1.

Auditory stimuli

For the eight experimental conditions auditory tones were identical to those in Experiment 1 (see Figure 1). Deviant trials were broadband noises with an intensity change of either ± 65% as opposed to the 15% change used in all other trials. All other parameters (e.g. duration, frequency range) were the same.

Audio-visual stimuli

The movie sequence was identical to Experiment 1 with two exceptions. First, audio clips in approximately 11% of trials were substituted with the deviant looming or receding sound. Second, fixation frame durations were quasi-randomly jittered at lengths of 2.6, 2.8, 3.1 and 3.4 s to limit anticipatory ERPs (the normalising vergence animation was maintained at 0.5 s). A practice movie with 12 trials including two deviant trials was created.

Procedure

Participants were required to press a button on a response pad when they identified the deviant sound. All participants received an initial training session of 12 trials to ensure they understood the task and could identify the deviant stimuli. The main experiment was then run as three blocks of 216 trials.

EEG data acquisition

EEG data was recorded from 64 electrodes using a BioSemi Active Two system (BioSemi, Amsterdam, Netherlands). Electrodes were placed according to the extended 10–20 system (Nuwer et al., 1998). Four additional leads were placed above and below the left eye and on the outer canthi of the left and right eyes, to record the vertical and horizontal electrooculogram (VEOG and HEOG, respectively). EEG signals from all channels were

acquired with respect to the common mode sense (CMS) electrode at a sampling rate of 512Hz.

ERP Analysis

The continuous EEG was divided into epochs offline, starting 100 ms prior to stimulus onset and ending 600 ms post-stimulus onset. The averages were digitally filtered (second-order zero-phase-lag bandpass filter, $1 - 25$ Hz). ERP amplitudes were aligned to a 100 ms pre-stimulus baseline period. ERPs were averaged according to audio-visual stimulus condition to produce eight ERPs per participant (see Figure 1).

EEG artefacts were rejected using the SCADS procedure with standard parameters (Junghöfer, Elbert, Tucker, & Rockstroh, 2000). This procedure initially detects artefacts for individual channels, then recomputes the data against the average reference and then detects global artefacts. Epochs that contained more than 10 unreliable sensors were excluded from analysis on the basis of the distribution of their amplitude, standard deviation and gradient. For the remaining epochs data from artefact-contaminated sensors was replaced by a statistically weighted spherical interpolation using all channels. With respect to the spatial distribution of the approximated electrodes, it was ensured that the rejected channels were not localised within one region of the scalp, because this would make interpolation for this area invalid. The standard deviation of the spherical splines used for approximation was computed for each epoch, and epochs containing outliers in this distribution were rejected. Across all participants and all conditions the procedure rejected an average of 30.1 % epochs as contaminated.

Statistical Analysis of ERPs

Global Field Power (GFP)
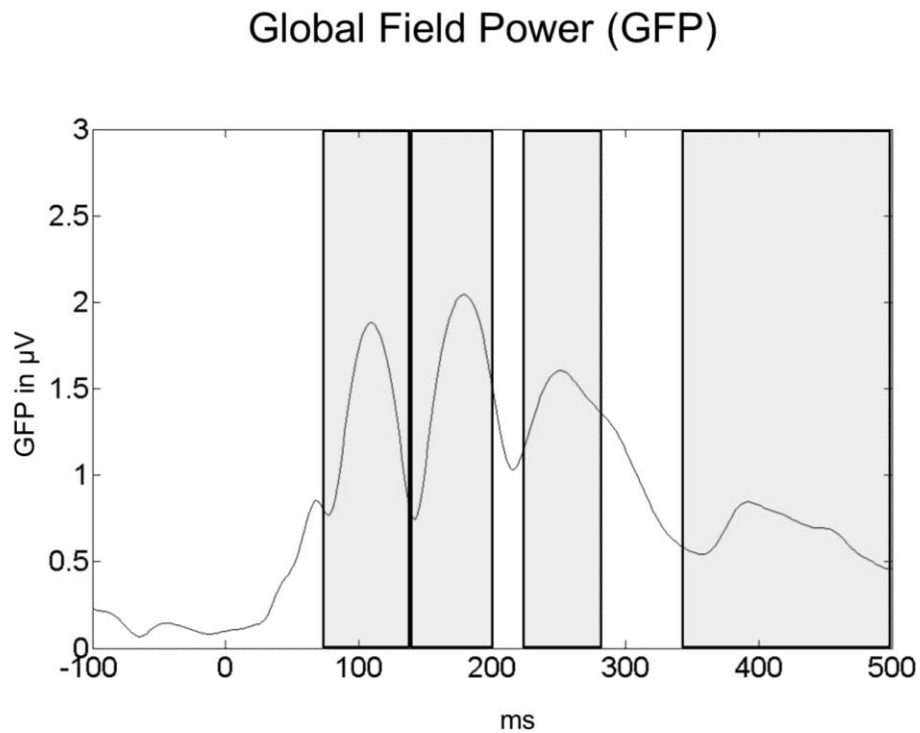
Fig. 3. *Global field power (GFP) for grand averaged waveforms, averaged across all stimulus conditions. Analysis windows for repeated measures ANOVA are marked in grey.*

In the first phase of analysis, ERPs were analysed during four time windows corresponding to the major peaks of the global field power (GFP) distribution (Figure 3), which also coincide with time-points previously identified as relevant in the review above. In each time window, amplitudes were averaged in two clusters composed of adjacent electrodes. Bilateral electrode clusters were selected to cover the area of maximal amplitude, as revealed by topographical mapping.
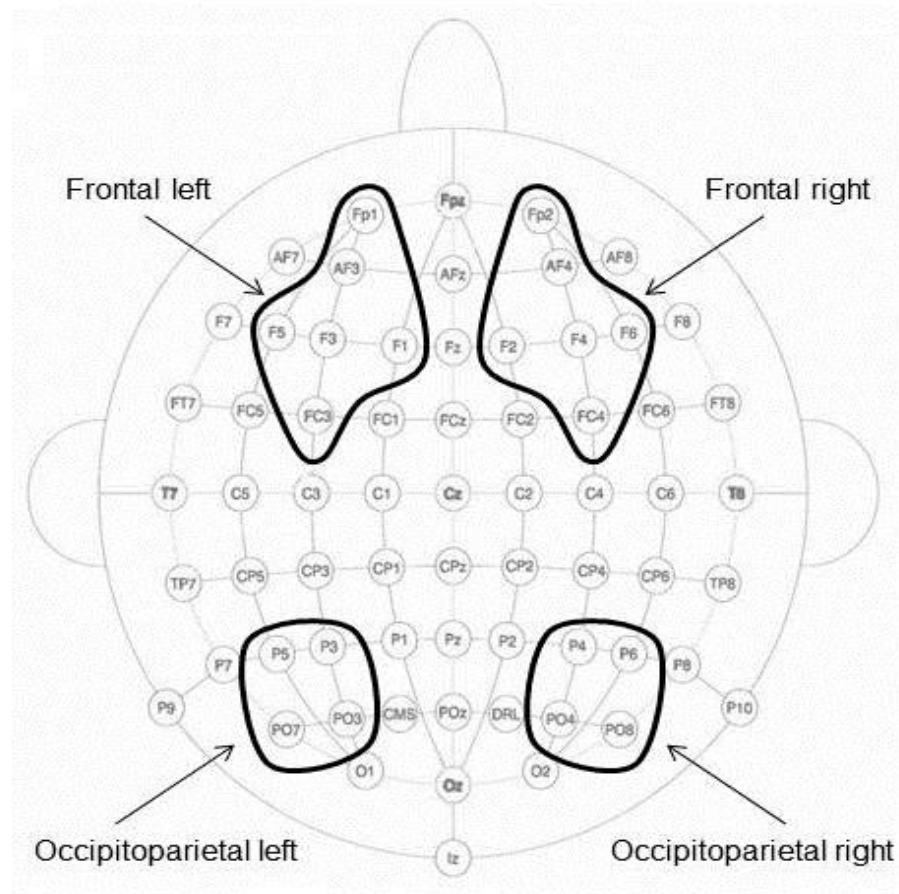
Fig. 4. *Top view of electrode positions showing the 64 EEG channels with the*

four clusters used in the analysis highlighted.

The first major peak of the GFP occurred around 110ms post-stimulus onset and activity was

maximal over bilateral occipito-parietal cortex, which closely corresponds to the latency and

scalp topography of the visual P1 component (e.g., Di Russo, Martinez, Sereno, Pitzalis, &

Hillyard, 2002). Amplitudes were calculated between 80 – 140 ms at left (electrodes: PO3,

PO7, P3, P5) and right (electrodes: PO4, PO8, P4, P6) occipito-parietal clusters (see Figure 4).

The second GFP peak occurred at around 170 ms post-stimulus, and was maximal over bilateral

occipito-parietal electrodes, corresponding to the visual N1 component (e.g., Clark, Fan, &

Hillyard, 1995; Vogel & Luck, 2000). Amplitudes were derived between 140 – 200 ms at the

same occipito-parietal electrodes clusters as described for the visual P1 component. Between

220 – 280 ms (the third peak in the GFP, with a peak latency of around 250 ms), analysis

focussed on frontal cortex (left cluster: electrodes FC3, F1, F3, F5, AF3, FP1; right cluster: FC4, F2, F4, F6, AF4, FP2), where the voltages were greatest. Peak amplitudes within each of the three time windows were analysed using repeated measures ANOVAs with the factors 'disparity' (2D, 3D), 'congruence' (congruent, incongruent), 'visual motion direction' (looming, receding), and 'scalp laterality' (left, right). For the fourth GFP peak at around 400 ms, ERPs were analysed at frontal electrodes during a longer time window between 350 – 500 ms, as the GFP wave exhibited a flatter profile at this latency. Mean amplitudes within this time window were analysed using a repeated measures ANOVA with the same four factors as for the first three peaks.

Effects of retinal disparity on audio-visual motion congruity

In this analysis phase, a more exploratory approach was adopted to investigate ERP waveforms and topographies in relation to effects of retinal disparity on the audio-visual motion congruity effect.

Firstly, a mass univariate analysis approach (e.g., Groppe, Urbach, & Kutas, 2011) was employed where all 64 electrodes were included in the statistical analysis for the full 500 ms following stimulus onset, to give a complete representation of differences between conditions with no a-priori assumptions about effect locations or latencies. This approach has been widely used in the literature on multisensory processing (see e.g., Besle et al., 2004; Giard & Peronnet, 1999; Meyer, Harrison, & Wuerger, 2013; Molholm et al., 2002; Vroomen & Stekelenburg, 2009). Specifically, we compared the congruity effect difference wave for 2D versus 3D presentation:

$$\boldsymbol{D} = (\boldsymbol{V_L A_R} + \boldsymbol{V_R A_L}) - (\boldsymbol{V_L A_L} + \boldsymbol{V_R A_R})$$

where $\boldsymbol{V}$ and $\boldsymbol{A}$ are visual and auditory motion in the looming (*L*) and receding (*R*) direction.

It is important to note that both sides of the difference wave contain exactly the same stimuli (V_L, A_L, V_R, and A_R) and that only the directional congruence between the signals was manipulated. We corrected for Type 1 error due to multiple comparisons by using the Guthrie and Buchwald (1991) procedure, where significance was defined as at least 12 consecutive time-points at a $p < .05$ level, at more than one adjacent electrode location.

Global topographic ERP analysis of retinal disparity and audio-visual motion congruity

In addition to the traditional ERP analysis we also performed analyses on two global variables, that is, the global field power (GFP) and the global dissimilarity (DISS) (Lehmann & Skrandies, 1980; Murray, Brunet, & Michel, 2008). The global topographic ERP pattern analyses represent a data-driven approach that aimed to assess response strength and topographic differences between the congruity effects for 2D and 3D viewing conditions.

Electric field strength analysis

To investigate the strength of the cortical response of the congruity effect in the 2D and 3D conditions we used the reference-independent Global Field Power (GFP, Lehmann & Skrandies, 1980; also see Fig. 3) measure. GFP is a measure of the scalp electric field strength, calculated as the standard deviation across all electrodes at a particular time point (Murray et al., 2008), and can be used to assess differences in the electric field strength of the EEG signal between conditions (Lehmann & Skrandies, 1980). To test for differences in electrical field strength between the 2D and 3D congruity conditions at successive time-points, a non-parametric randomization test was conducted on the GFP between the 2D and 3D congruity conditions (Koenig, Kottlow, Stein, Melie-García, 2011). The GFP was the dependent variable in the randomization-based analysis, calculated from the mean ERPs of the 2D and 3D congruity effects for each participant. To control for multiple comparisons in the randomization-based analysis (and in the TANOVA analysis reported below), we

considered as significant an alpha level of *p* less than 0.05 at at least three consecutive time points (c.f., Maurer, Rossion, & McCandliss, 2008). The joint probability of $p < 0.05$ at three successive time points ($0.05 \times 0.05 \times 0.05$) exceeds the Bonferroni corrected threshold of p < 0.05 across the 307 time points tested in the randomization test (0.05/307).

GFP latency analysis

To test for latency shifts in the strength of the cortical response as indexed by the GFP, peak latencies were analysed on GFP values over all electrodes (Hauk & Pulvermuller, 2005). Peak latencies were determined in each condition for each subject, and the peak latency was defined as the time point at which the GFP reached a maximum within the 100 – 160 ms time window. These values were compared for 2D and 3D congruency effects using a two-tailed paired t-test.

Electric field topography analysis

Here we tested for topographic differences of the scalp field map between the 2D and 3D congruity effects, independent of the strength of the response (i.e., independent of the GFP). Topographic differences independent of GFP can be quantified using a measure known as global dissimilarity (DISS) (Lehmann & Skrandies, 1980; Murray, Brunet, & Michel, 2008). After first normalized the ERP data by GFP at each time point, the square root of the mean of the squared differences between the normalized scalp fields at each time point was statistically analysed by a so-called topographic ANOVA (TANOVA) randomization test (Murray, Brunet, & Michel, 2008). The topographic maps of single participants were randomly reassigned to either the 2D or 3D congruity effect condition, and the global dissimilarity (DISS) index for the randomly permutated data was compared at each time point with the global dissimilarity index of the actual conditions. A randomisation test with 2000

runs was conducted using customised scripts in MATLAB, and corrections for multiple comparisons were the same as described above for the GFP randomisation test.

Results

Behavioural: Deviant trial detection task

For the EEG experiment an irrelevant 'oddball' task was employed to promote attention to the experimental stimuli while minimising manual responses. The task involved detection of a deviant looming or receding sound, quasi-randomly interspersed across trials. Participants correctly identified 77% of the target signals (chance performance level = 11%), while false positives responses were given in only 0.06% of trials. The data confirms that participants attended to the stimuli and that the task difficulty was acceptable.

ERPs

Grand averaged waveforms showed a typical series of components for audio-visual potentials, consisting of a summation of visual and auditory responses (Meyer, Harrison & Wuerger, 2013; Proctor & Meyer, 2012).

80- 140 ms (Visual P1)

ERP amplitude was maximal over bilateral occipito-parietal cortex at around 110 ms post-stimulus, corresponding to the latency and scalp topography of the visual P1 component. Peak amplitude between 80 – 140 ms was analysed at left (electrodes: PO3, PO7, P3, P5) and right (electrodes: PO4, PO8, P4, P6) occipito-parietal clusters. A four-way repeated measures ANOVA with factors Congruency, Motion Direction, Disparity, and Laterality revealed no significant main effects or interactions (all $ps > .21$).

140 – 200 ms (Visual N1)

N1
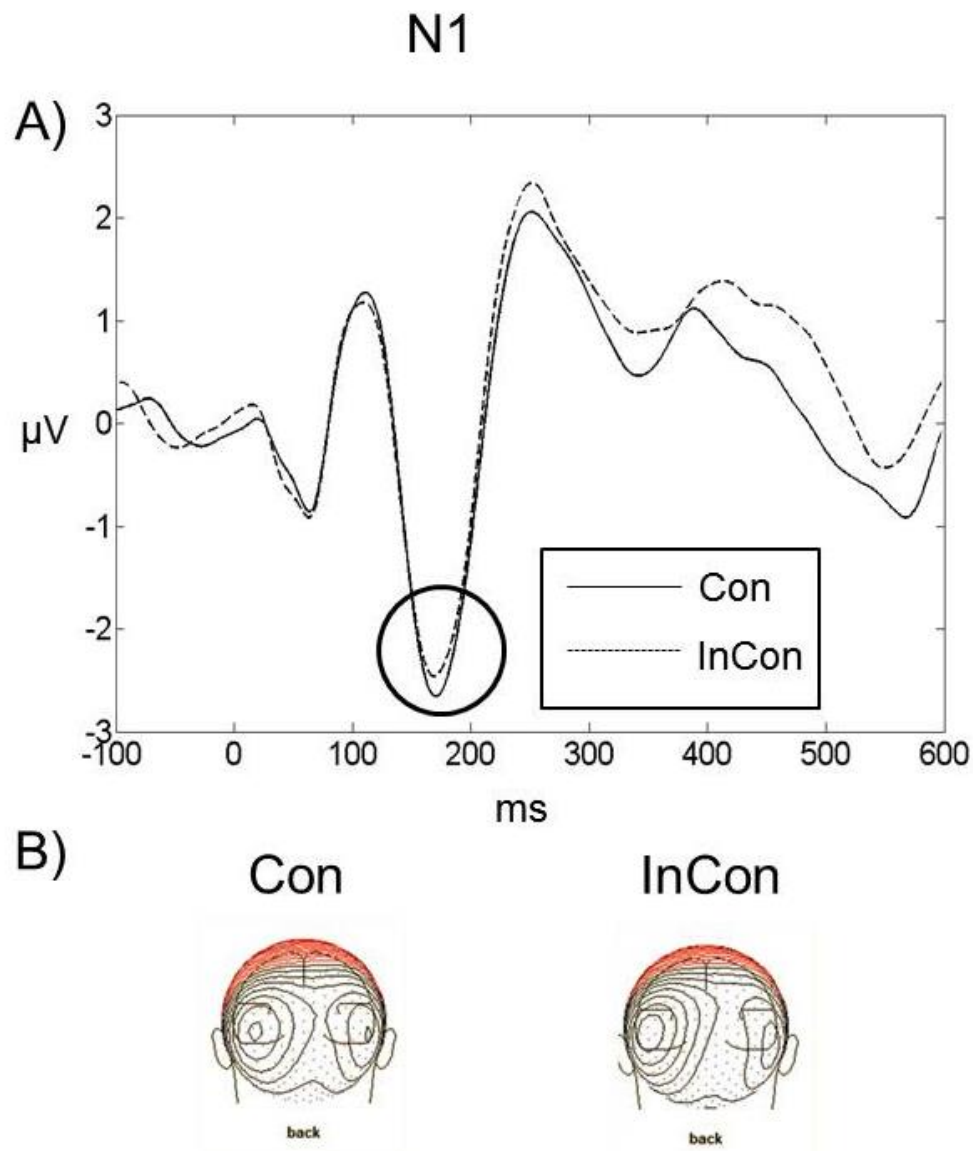
Fig. 5. *A) The left occipito-parietal electrode cluster amplitude for the congruent condition was more negative than incongruent between 140 – 200 ms. B) Scalp maps show that activation at 170 ms is maximal over bilateral extrastriate visual areas. Red lines indicate positive amplitude, black lines indicate negative amplitude.*

At around 170 ms, ERP amplitude was maximal over bilateral occipito-parietal cortex, corresponding to the latency and scalp topography of the visual N1 component (see Figure 5).

Between 140 – 200 ms peak amplitudes were calculated at the same occipito-parietal electrodes clusters as for the visual P1 component. A four-way repeated measures ANOVA revealed a significant interaction between congruency and laterality ($F(1,10) = 7.68$, $p = 0.020$). Post-hoc paired t-tests found a significant difference between congruent and incongruent presentation over left occipito-parietal cortex, where congruent amplitudes were more negative than incongruent ($t (10) = 2.657$, $p = 0.024$; mean amplitude for congruent = -2.88, SD = 1.68; mean amplitude for incongruent = -2.48, SD = 1.44). There was no difference over right occipito-parietal cortex ($t (10) = 0.368$, $p = 0.721$). The four-way repeated measures ANOVA also revealed a significant main effect of disparity ($F(1,10) = 6.02$, $p = 0.034$) where the amplitude was more negative for 2D compared to 3D presentation.

220 – 280 ms (N2)

N2

A)

B)
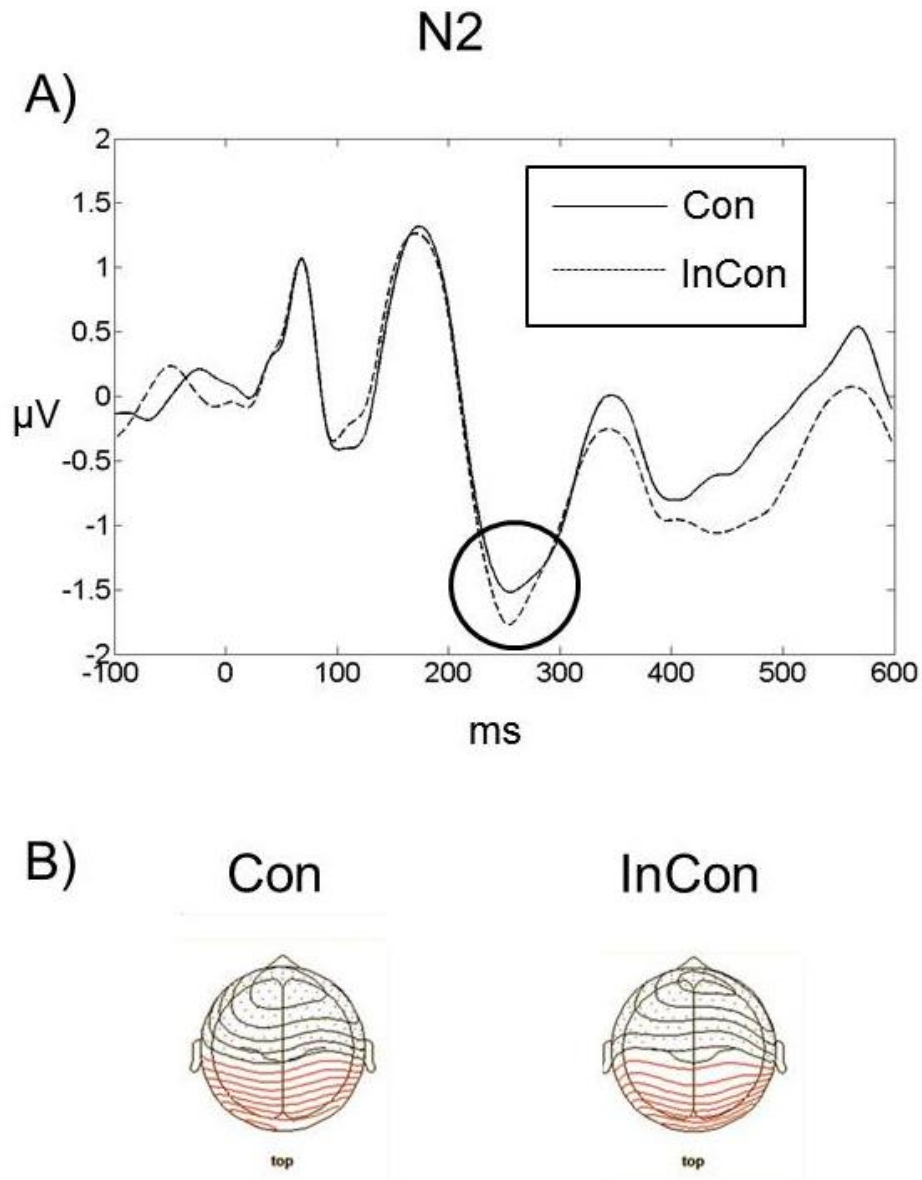
Fig. 6. A) The amplitude for the incongruent condition was more negative than for the congruent condition over both frontal electrode clusters between 220 – 280 ms. B) At 250 ms the greatest activity is over frontal scalp.

At around 250 ms post-stimulus, there was a large negative deflection over frontal leads (see Figure 6). Activity at this latency and topography most likely corresponds to the N2 component, which is thought to be related to conflict monitoring processes (Folstein & Van

Petten, 2008). Mean amplitudes between 220 – 280 ms at electrode clusters over left (electrodes FC3, F1, F3, F5, AF3, FP1) and right (electrodes FC4, F2, F4, F6, AF4, FP2) frontal scalp were submitted to a four way repeated measures ANOVA. We observed a main effect of congruency ($F(1,10) = 6.13$, $p = .033$), where incongruent amplitudes were more negative than congruent amplitudes.

350 – 500 ms (N400)
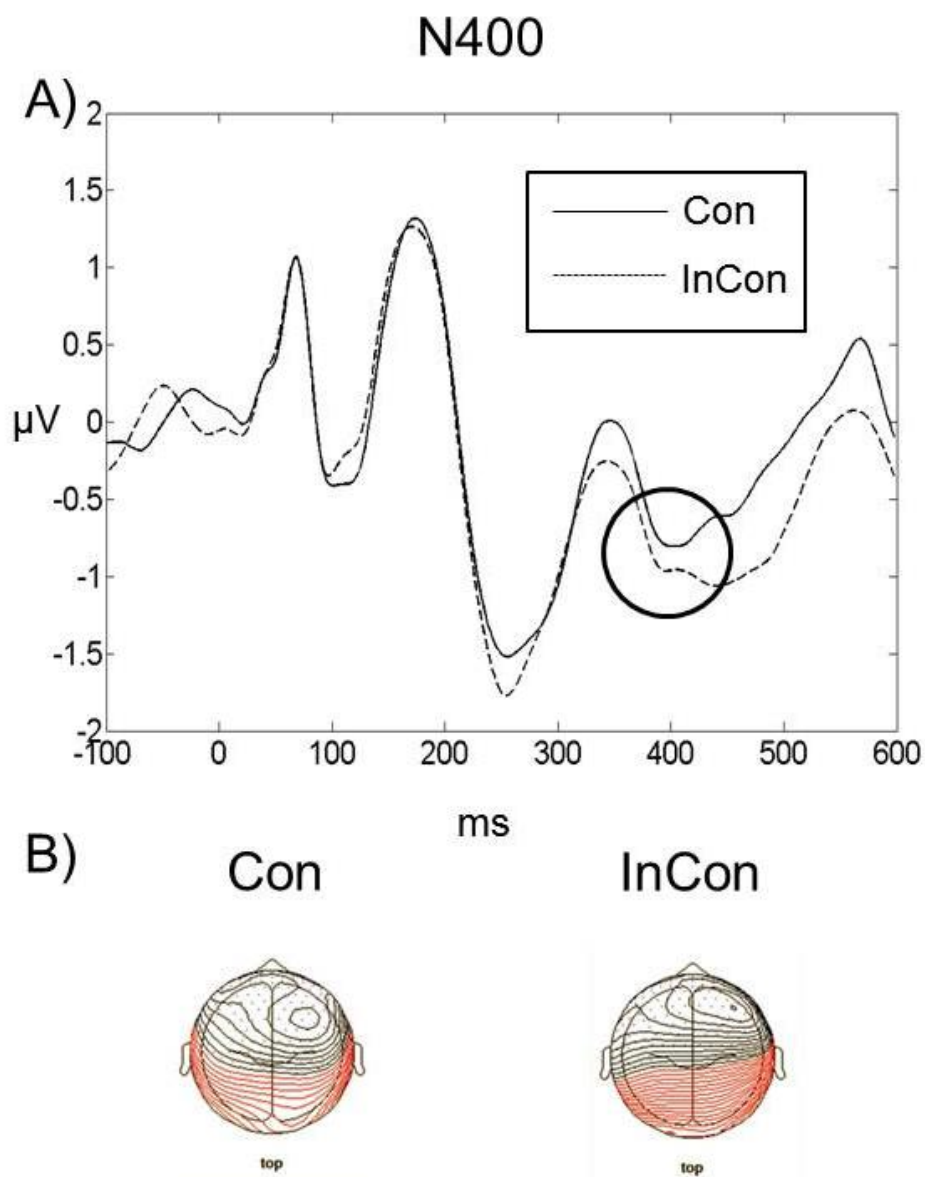
Fig. *7.* *A) The amplitude for incongruent is more negative than congruent between 350 – 500 ms over both frontal electrode clusters. B) At 400 ms the greatest activity is over right frontal scalp*
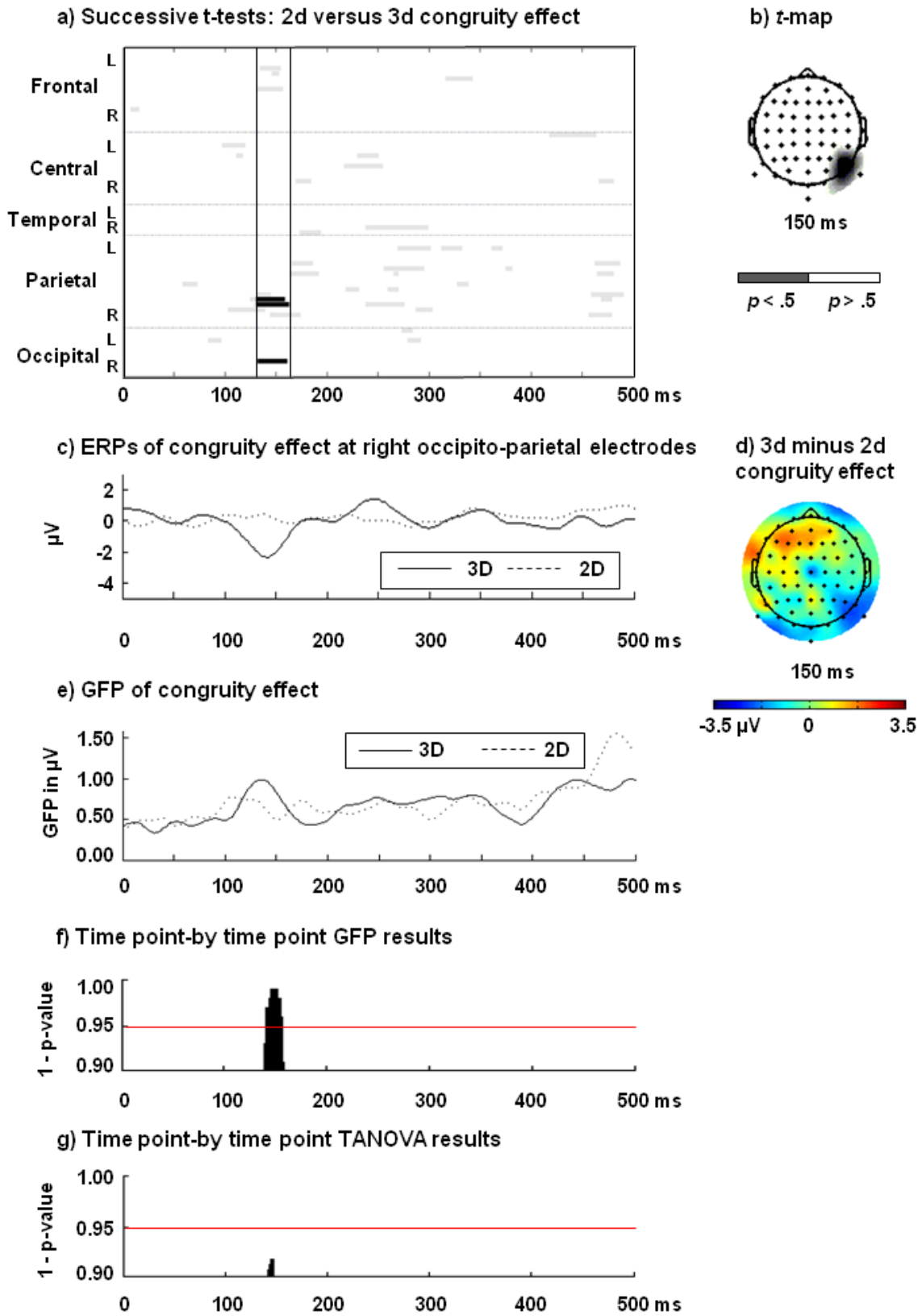
Between around 350 – 500 ms post-stimulus there was a negative deflection which was maximal over frontal electrodes. Left and right frontal electrode clusters comprising the same electrodes as for the N2 component described above were subjected to a four way repeated-measures ANOVA, which revealed a significant main effect of congruence ($F(1,10) = 15.50$, $p = 0.003$), where incongruent amplitudes were more negative than congruent amplitudes (see Figure 7). This main effect was moderated by a marginally significant three-way interaction between congruency, looming and laterality ($F (1,10) = 4.49$, $p = 0.06$). A simple interaction effects analysis was conducted using a two-way ANOVA with factors congruency and region at each level of the factor looming, and this revealed a significant interaction between congruency and region for looming signals ($F(1,10) = 5.91$, $p = 0.03$), and post-hoc paired t-tests showed that amplitudes were more negative ($t(10) = 2.277$, $p = 0.046$) in the incongruent (mean = -.79, SD = .99) compared to congruent (mean = -0.40, SD = .61) condition for the frontal right region, but not for the frontal left region ($t(10) = 0.14$, $p = 0.888$). There were no main effects or interactions for receding signals (all ps > .3).

Effects of retinal disparity on audio-visual motion congruity

In this analysis step, data-driven approaches were used to investigate ERPs waveforms and scalp topographies in relation to retinal disparity cues. The aim of this comparison was to evaluate the effect of retinal disparity on the audio-visual motion congruity effect (i.e., discrepancy between congruent and incongruent conditions). We contrasted the 2D congruency difference wave (i.e., 2D incongruent minus 2D congruent) with the 3D

congruency difference wave (i.e., 3D incongruent minus 3D congruent), to assess whether the congruity effect was stronger in 3D or 2D.

Firstly we used a mass-univariate analysis to statistically compare (using paired t-tests at each time-point) the 2D and the 3D congruency difference waves (see Fig. 8a & c). The comparison clearly showed that the congruency effect was modulated by disparity (2D vs. 3D) between 135 to 160 ms after stimulus onset, over right occipito-parietal scalp (see Fig. 8b & d). The congruity effect was stronger in 3D presentation than 2D presentation, and the results indicate that retinal disparity cues affected the congruity effect at a perceptual stage of processing.

a) Successive t-tests: 2d versus 3d congruity effect

b) t-map

150 ms

$p < .5$ $p > .5$

c) ERPs of congruity effect at right occipito-parietal electrodes

3D ------ 2D

d) 3d minus 2d congruity effect

150 ms

-3.5 µV   0   3.5

e) GFP of congruity effect

3D ------ 2D

f) Time point-by time point GFP results

g) Time point-by time point TANOVA results

Fig. 8. *Results of the ERP and topographic analyses of the 2D versus 3D congruity effects.*
(a) Significant p-values for all 64 electrodes showing differences in congruency effects (i.e.,
incongruent minus congruent) between 2D and 3D presentation. Bold lines indicate periods
of significant p-values after correction for multiple comparisons (Guthrie & Buchwald,
1991). Significant differences between 2D and 3D presentation emerged between 135 – 160
ms after stimulus onset. Electrode locations are plotted on the y-axis, and time following
stimulus onset on the x-axis. (b) A scalp plot of the locations of the significant differences in
congruity effect between 2D and 3D at 150 ms, showing that the congruity effects differed
over right occipito-parietal scalp. (c) Plot of the 2D (dashed line) and 3D (solid line)
congruency difference waves, showing that the congruity effect was stronger in 3D than 2D
between 135 – 160 ms post-stimulus. (d) Scalp plot of the 3D – 2D congruity effect at 150 ms.
(e) Cortical response strength as indexed by global field power (GFP) is plotted for the 2D
(dashed line) and 3D (solid line) congruency difference waveforms. (f) Statistical results of
the time point-by time point randomisation test for GFP differences between 2D and 3D
congruity effects. Significant differences were observed between 142 – 158 ms after stimulus
onset. g) Results of the TANOVA analysis, which indicated an absence of topographic
differences between 2D and 3D congruity effects.

Next we conducted a global topographic ERP analysis to assess the effects of retinal disparity
on the audio-visual motion congruity effect in relation to cortical response strength and scalp
distribution. The effects of 2D and 3D presentation on the cortical strength of the congruity
effects as measured by the GFP were analysed using a statistical randomization test (Koenig,
Kottlow, Stein, Melie-García, 2011) for latencies between 0 – 500 ms after stimulus onset.
The difference between 2D and 3D response strength (GFP) reached significance between

142 – 158 ms (see Fig. 8f), where response strength was greater for 3D compared to 2D presentation.

A time-point by time-point TANOVA (Murray, Brunet, & Michel, 2008) was conducted to evaluate whether there were systematic differences in scalp topography, independent of response magnitude, between 2D and 3D congruity conditions. Results of the randomisation test are presented in Fig. 8g, where it can be seen that there was no evidence of statistically significant topographic differences between 2D and 3D congruity effects. GFP differences in the absence of topographic modulations are most parsimoniously interpreted as differential activation strength of a common brain network (Murray et al., 2008).

Lastly, we tested whether the modulation of response strength (as represented by GFP) between 2D and 3D congruity effects was related to latency differences in GFP peaks between 100 – 160 ms (see Fig. 8e). We found no evidence of GFP latency shifts (t(10) = 0.766, *p* = .462; 2D peak latency: mean = 131.89 ms, SD = 22.44; 3D peak latency: mean = 140.41 ms, SD = 19.85).

General discussion

We investigated the effect of the directional congruence of audio-visual stimuli on the perception of motion in depth, and the modulation of this effect by motion direction and the retinal disparity of the visual cues. In Experiment 1, we found enhanced speed and accuracy for auditory motion detection in congruent relative to incongruent audio-visual trials. Importantly, response accuracy was mediated by visual motion direction and stereo disparity. In Experiment 2, we used ERPs to assess the time-course of differences between congruent and incongruent audio-visual motion directions. Our main finding from Experiment 2 was that the congruity effect (difference between incongruent and congruent conditions) was affected by retinal disparity at an early processing stage (135 – 160 ms), where 3D

presentation showed an enhanced congruity effect compared to 2D presentation. Global topographic analyses sshow that the cortical response strength was enhanced in the 3D compared to the 2D congruity effect, but that the scalp topographies did not differ between these conditions. Differences between directionally congruent and incongruent audio-visual stimuli also emerged between 140 – 200 ms, 220 – 280 ms, and 350 – 500 ms, reflecting a series of stages when neural processing related to audio-visual motion in the depth plane is accomplished.

Together, the results of Experiment 1 are in line with studies showing visual modulation of audio motion in depth perception (Harrison, 2012; Jain et al., 2008). Similarly, enhanced behavioural responses for looming visual stimuli have previously been reported in unimodal (Ball & Tronick, 1971; Schiff et al., 1962) and multimodal (Cappé et al., 2009; Harrison, 2012) paradigms, and have been interpreted to be a result of the adaptive salience of approaching objects. The reaction times we observed were significantly slower than those reported by Cappe et al. (2009) who reported response times for the detection of (any) looming motion in congruent audio-visual signals between 400 and 500ms. Brooks et al. (2007), similar to our behavioural experiment, asked participants to discriminate the direction of biological *lateral* motion and reported reaction times ranging from 1.2s to 2.2s that were strongly modulated by task difficulty and the congruence of audio-visual signals. The strongest modulation of response times by congruency were observed in the most difficult task settings. Our response times, ranging from 1.05s (congruent) to 1.2s (incongruent) were faster than the those reported by Brooks et al. (2007).

Our results for 3D cue presentation contrasts with the behavioural findings of Ogowa and Macaluso (2013), and instead supports claims about the involvement of disparity cues in collision avoidance (Tresilian, Mon-Williams, & Kelly, 1999). However, differences between disparity conditions were relatively small, adding to arguments claiming their relative

subordination compared to other cues, such as visual expansion (Yan, Lorv, Li, & Sun, 2011).

Experiment 2 used event-related potentials (ERPs) to investigate the timing of neural processing associated with the perception of audio-visual motion in depth, while participants were engaged in a deviant-trial detection task. Our analysis focussed on the timing of the crossmodal congruity effects i.e., on differences between congruent conditions (where visual and auditory signals moved in the same direction) and incongruent conditions (where visual and auditory cues moved in opposite directions), and in particular we were interested in the influence of effects of retinal disparity on the congruity effects.

Congruency as a measure of bimodal interaction has been proposed as a method for assessing multisensory processes that avoids some potential confounds in the comparison of bimodal ERPs (Gondan, Niederhaus, Rösler, & Röder, 2005; Proctor & Meyer, 2011; Meyer, Harrison, & Wuerger, 2013). With this method, bimodal stimuli are either congruent or incongruent on one dimension (in our case, motion direction), so activity common to both conditions (for example, motor activity) is eliminated from the analysis (Calvert & Thesen, 2004). This method has successfully been used to investigate crossmodal semantic (Meyer, Harrison, & Wuerger, 2013) and spatial congruity (Gondan, Niederhaus, Rösler, & Röder, 2005), as well as multisensory perception of faces (Proctor & Meyer, 2011).

It provides an alternative to the more commonly used analysis of crossmodal interactions that is based on the subtraction method (the so-called 'additive model'; Besle, Fort, & Giard, 2004; Giard & Besle, 2010), where ERPs time-locked to unimodal stimuli are subtracted from the bimodal responses (e.g. Giard & Perronet, 1999; Cappe et al., 2012). This model has been criticised because results may be affected by the unequal subtraction of common activity (e.g., anticipatory slow wave potentials and task-irrelevant motor activity) from the

AV and (A + V) amplitudes, and artefacts caused by inequivalent attentional demands in the unimodal and bimodal conditions (Giard & Besle, 2010; Teder-Sälejärvi, McDonald, DiRusso, & Hillyard, 2002). There are various approaches that employ the additive model, particularly in fMRI analysis, that are not associated with these confounds (e.g. Werner & Noppeney, 2010; Saldern, & Noppeney, 2013).

Our most noteworthy ERP result was that motion direction congruity was influenced by retinal disparity starting at around 135 ms post-stimulus onset, which is a relatively early stage of processing. This ERP effect was observed over right occipito-parietal scalp (Figure 8b & d), where 3D presentation was associated with an increased congruity effect compared to 2D presentation, suggesting that retinal disparity facilitated the discrimination of directional discrepancy between the cues. Importantly, the 2D versus 3D congruity effect cannot be explained by physical differences (other than retinal disparity) between conditions, as exactly the same stimuli were present on each side of the difference wave equations. The latency of the effect (135 – 160 ms) occurs somewhat prior to the time range of the visual N1 component (150 – 200 ms) suggesting that the enhanced congruity effect for 3D presentation does not reflect a simple enhancement of the visual N1. Global analyses show that the 2D versus 3D congruity effect was manifested as an increase in cortical response strength in the 3D condition but that the scalp topography of the 2D and 3D conditions did not differ. The most plausible explanation for this pattern of results, therefore, is that 2D and 3D cues engage a similar cortical network (cf. Murray et al., 2008), but that neural activation in the network is greater under 3D viewing conditions. The right occipito-parietal scalp topography of the 3D congruity effect is consistent with studies showing that disparity cues are integrated at an early stage in the dorsal visual stream when the signals are moving (e.g., Howard, 2012). The effect of stereo disparity in our study occurred slightly earlier than in a previous study that showed differences between mono and stereo cues at around 170 ms over occipito-temporal

leads (Pegna, Roberts, & Leek, 2014). Task differences between the two experiments could potentially explain the apparent discrepancy; Pegna and colleagues used an object recognition task with stationary visual cues, whereas in the current study we used an auditory detection task with dynamic cues. These observations suggest that further studies are needed to more fully elucidate how the timing of ERPs related to stereo disparity are modulated by task requirements and stimulus conditions (e.g., moving versus stationary cues).

Visual disparity cues are the basis of stereopsis, the perception of depth from differences in retinal images in the two eyes. A commonly held view (e.g., Barry, 2009) is that the processes underlying stereopsis are generally slow. Rapidly changing disparities, for example, are perceptually difficult to track and stereoacuity improves with exposure duration. The temporal resolution of the stereoscopic system for stimuli that fluctuate in depth is about 10 Hz, as compared with 70 Hz for luminance modulation, which has been explained by the requirement to cross-correlate temporally filtered inputs from the two eyes (Kane, Guan, & Banks, 2014; Nienborg, Bridge, Parker, & Cumming, 2005).

In one of the earliest experiments, Langlands (1929), reported increasing stereoacuity with exposure durations of up to 3 seconds, while more recently Watt (1987) showed a linear relationship between disparity threshold and exposure time up to about 1 sec. Tyler (1991), consistently, measured the disparity required for the detection of depth in a random-dot stereogram and found that thresholds were inversely proportional to stimulus duration, such that fine disparities (<1 arcmin) required around 180ms to be detectable.

Disparity cues, in summary, appear to provide accurate distance information that requires time to compute. It has been argued that this characteristic limits the use of stereopsis to conscious appreciation of depth or actions that can be planned ahead of time (e.g., Keefe,

Hibbard, & Watt, 2011; Schlicht & Schrater, 2007), or slow tasks such as threading a needle (Brenner & Smeets, 2006; Sheedy, Bailey, Buri, & Bass, 1986).

This, however, does not exclude a role of (inaccurate) disparity information in early visual processing. The nervous system extracts disparity very quickly, with disparity-selective responses evident in cortical neurons around 60 ms after stimulus presentation in macaque monkeys, which is similar to other visual features such as orientation (Trotter, Celebrini, Stricanne, Thorpe, & Imbert, 1996). If binocular disparity is processed at early visual processing stages and at time-scales that are consistent with other visual features, then there is no reason to think that disparity information should not contribute to everyday visual function, particularly those that require rapid decisions, perhaps based on inaccurate incoming data.

Caziot et al. (2015) presented experimental data comparing speed–accuracy trade-off functions between 2 forced-choice discriminations: one based on stereoscopic depth, the other based on luminance. Both speed-accuracy trade-off functions deviated from chance levels of accuracy at the same, early, response time (200 ms) with stereo accuracy increasing, on average, more slowly than luminance accuracy after this initial delay. This timescale is consistent with electrophysiological data from nonhuman primates that shows involvement of primary visual cortex at simultaneously with other visual processing (Gonzalez, Perez, Justo, & Bermudez, 2001; Trotter et al., 1996). The task is not unlike our motion judgement task where participants could use expansion and disparity cues to judge motion direction. We see electrophysiological correlates of disparity information at around 150ms (Fig. 8) which, is consistent with the effects reported by Caziot et al. (2015).

A temporal dependence of judgement accuracy on the basis of disparity cues may also explain why the addition of disparity cues in the behavioural task lead to significant improvements in response accuracy but only weakly affected response times.

In our canonical ERP component analysis, we analysed directional congruity effects for four specific ERP components (visual P1, visual N1, N2, N400) evoked by the audio-visual stimuli. For the visual P1 component (80 – 140 ms) over occipito-temporal scalp, we found no effect of congruity. A number of previous studies have reported multisensory effects at latencies around 100 ms for stationary signals (e.g., Besle et al.,2004; Giard & Peronnet, 1999; Teder-Sälejärvi et al., 2002, 2005), and it may be that multisensory integration for moving signals is delayed compared to interactions for stationary cues. On the other hand, Cappe and colleagues (2012) found multisensory interactions for looming (ALVL) and incongruent ALVR audio-visual motion cues at a similar latency, although their topographical analysis suggested that the generators were located more anteriorly, in the claustrum/insula and cuneus. A number of key differences between Cappe et al.'s study and the current study analysis may explain the apparent discrepancy. In particular, Cappe et al., focused on differences between multisensory and unisensory presentation in a motion detection task, whereas in the current study only bimodal stimuli were presented, and the task was to detect deviant trials.

At the latency and topography of the visual N1 (140 – 200 ms), we observed more negative amplitudes in the congruent conditions than in the incongruent conditions over left occipito-parietal leads (Figure 5). Modulation of the visual N1 in response to multisensory inputs has previously been reported (e.g., Giard & Perronet, 1999) using stationary stimuli, but here we show crossmodal modulation of the visual N1 specific to motion in depth. Our data reinforces the view that processing in extrastriate visual areas, as reflected by the visual N1 (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002), can be influenced by auditory motion inputs. It

is worth noting that a previous study investigating crossmodal semantic rather than low-level congruence did not show congruency effects for the visual N1; instead the congruency effects emerged later, presumably as a result of their higher-order (semantic) nature (Meyer et al., 2013).

Between 220 – 280 ms post-stimulus onset incongruent ERPs were more negative than congruent ERPs over frontal scalp (Figure 6). The prominent negative deflection at this latency is likely to be related to the N2 component, which is thought to reflect conflict monitoring processes (Folstein & Van Petten, 2008; Yeung, Botvinick, & Cohen, 2004). While N2 has been interpreted as a result of response conflict (Yeung, Botvinick, & Cohen, 2004), the current results are in keeping with previous crossmodal studies that have shown an increased N2 response for visuo-tactile incongruent compared to congruent trials (Forster & Pavone, 2008; Longo, Musial, & Haggard, 2012). Indeed, given that a deviant trial detection task was used, the current study strongly suggests that the increased N2 for incongruent trials reflected perceptual conflict between the auditory and visual pairing, rather than processes related to response conflict.

At around 400ms, we observed more negative amplitudes in the incongruent condition compared to the congruent condition for bimodal pairings containing visual looming signals (Figure 7). The latency of this effect in our study may well reflect an N400-like effect, which is in general agreement with several previous studies showing N400-like enhanced negativity for semantically incongruous AV signals (e.g. Meyer, Harrison, & Wuerger, 2013; Proctor & Meyer, 2011; Zimmer et al., 2010) and affectively incongruous AV stimuli (Goerlich et al., 2012). While the N400 is generally associated with incongruences in linguistic stimuli (for a review, see Kutas & Federmeier, 2011), the N400 has also been shown to be evoked by non-linguistic incongruency related to, for example, line drawing and sounds (Cummings, Čeponienė, Koyama, Saygin, Townsend, & Dick, 2006; Ganis, Kutas, & Sereno, 1996).

Moreover, the linguistic N400 component is usually reported as a centro-parieto deflection (Kutas & Federmeier, 2011), whereas the current data indicates a more frontal effect, in line with the topography from studies on the non-linguistic N400 (Cummings, Čeponienė, Koyama, Saygin, Townsend, & Dick, 2006; Ganis, Kutas, & Sereno, 1996).

In summary, by investigating behavioural and electrophysiological correlates of audio-visual motion in depth, the current studies have demonstrated that the dynamic visual capture effect is mediated by both the direction of visual motion and the retinal disparity of the visual cues. Motion direction congruency between audio-visual cues appeared as a robust effect in the ERP analysis, and the congruity effect was mediated by retinal disparity as early as 135 ms after stimulus onset.

References

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology, 74*(1), 28-33.

Ball, W., & Tronick, E. (1971). Infant responses to impending collision - optical and real. *Science, 171*, 818–820.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature Neuroscience, 15*(4), 636-643.

Barry, S. R. (2009). Fixing my gaze: A scientist's journey into seeing in three dimensions. New York, NY: Basic Books.

Besle, J., Bertrand, O., & Giard, M. H. (2009). Electrophysiological (EEG, sEEG, MEG) evidence for multiple audio-visual interactions in the human auditory cortex. *Hearing Research, 258*(1-2), 143.

Besle, J., Fort, A., & Giard, M. H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing, 5*, 189-192.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225-2234.

Billington, J., Wilkie, R. M., Field, D.T., & Wann, J. P. (2011). Neural processing of imminent collision in humans. *Proceedings of the Royal Society B: Biological Sciences, 278*(1711), 1476-1481.

Bland, J. M., & Altman, D. G. (1996). Transformations, means, and confidence intervals. *British Medical Journal, 312*(7038), 1079.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America, 102*(51), 18751-18756.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology, 17*(19), 1697-1703.

Brenner, E., & Smeets, J. B. J. (2006). Two eyes in action. Experimental Brain Research, 170, 302–311.

Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., & Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia*, *45*(3), 523-530.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*(1), 191-205.

Cappe, C., Thut, G., Romei, V., & Murray, M.M. (2009). Selective integration of auditory–visual looming cues by humans. *Neuropsychologia 47*:1045–1052.

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. M. (2012). Looming signals reveal synergistic principles of multisensory integration. *The Journal of Neuroscience, 32*(4), 1171-1182.

Caziot, B., Valsecchi, M., Gegenfurtner, K. R., & Backus, B. T. (2015). Fast perception of binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 909.

Clark, V. P., Fan, S., & Hillyard, S. A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping, 2*, 170–187.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A.P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research, 1115*, 92-107.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*, 95–111.

Erkelens, C.J., & van Ee, R. (1997). Capture of visual direction: An unexpected phenomenon in binocular vision, *Vision Research, 37*, 1193–1196.

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

Folstein, J.R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*, 152-170.

Forster, B., & Pavone, E.F. (2008). Electrophysiological correlates of crossmodal visual distractor congruency effects: Evidence for response conflict. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 65-73.

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12(*10), 1031-1039.

Franconeri, S. L. & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999-1010.

Ganis, G., Kutas, M., & Sereno, M.I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*, 89-106.

Giard, M.H., & Besle, J. (2010). Methodological considerations: Electrophysiology of multisensory interactions in humans. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory Object Perception in the Primate Brain* (55–70). New York: Springer.

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

Goerlich, K.S., Witteman, J., Schiller, N.O., Van Heuven, V.J., Aleman, A., & Martens. S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24*, 1725–1741.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception & Psychophysics, 67*(4), 713-726.

González, E. G., Allison, R. S., Ono, H., & Vinnikov, M. (2010). Cue conflict between disparity change and looming in the perception of motion in depth. *Vision Research, 50*(2), 136.

Gonzalez, F., Perez, R., Justo, M. S., & Bermudez, M. A. (2001). Response latencies to visual stimulation and disparity sensitivity in single cells of the awake Macaca mulatta visual cortex. Neuroscience Letters, 299, 41– 44.

Groppe, D.M., Urbach, T.P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*, 1711-1725.

Harrison, N. (2012). Auditory motion in depth is preferentially captured by visual looming signals. *Seeing and Perceiving, 25*(1), 71-85.

Hofbauer, M., Wuerger, S. M., Meyer, G. F., Roehrbein, F., Schill, K., & Zetzsche, C. (2004). Catching audiovisual mice: Predicting the arrival time of auditory-visual motion signals. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 241-250.

Howard, I. P. (2012). *Perceiving in Depth, Volume 1: Basic Mechanisms*. Oxford: Oxford University Press.

Howard, I. P., & Rogers, B. J. (2002). Seeing in depth: Vol. 2: Depth perception. Thornhill, Ontario, Canada: I. Porteous.

Jain, A., Sally, S.L., & Papathomas, T.V. (2008). Audio-visual short-term influences and aftereffects in motion: Examination across three sets of directional pairings. *Journal of Vision*, 8, 1-13.

Junghöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology, 37*(4), 523-532.

Kane, D., Guan, P., & Banks, M. S. (2014). The limits of human stereopsis in space and time. Journal of Neuroscience, 34, 1397–1408.

Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: No evidence for a binocular specialism. Neuropsychologia, 49, 1246 –1257.

Kim, J., Yang, H. K., Kim, Y., Lee, B., & Hwang, J. M. (2011). Distance stereotest using a 3-dimensional monitor for adult subjects. *American Journal of Ophthalmology, 151*(6), 1081-1086.

Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience*, 938925.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621-647.

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience, 1*(1), 38-49.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*, 609–621.

Longo, M.R., Musil, J.J., & Haggard, P. (2012). Visuo-tactile integration in personal space. *Journal of Cognitive Neuroscience, 24*, 543-52.

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: Apparent auditory motion induced by a moving visual target. *Perception, 14*(6), 721-727.

Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport*, *12*(11), 2557-2560.

Meyer, G. F., Wuerger, S. M., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, *166*(3-4), 538-547.

Meyer, G.F., Harrison, N.R., & Wuerger, S.M. (2013). The time course of auditory–visual processing of speech and body actions: Evidence for the simultaneous activation of an extended neural network for semantic processing. *Neuropsychologia,* 51 (9), 1716-1725.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research,* 14(1), 115-128.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step by-step tutorial review. *Brain Topography, 20*, 249–264.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2005). Neuronal computation of disparity in V1 limits temporal resolution for detecting disparity modulation. Journal of Neuroscience, 25, 10207– 10219.

Nuwer, M.R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J.M., Hinrichs, H., Ikeda, A., Luccas, F.J., & Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology (Supplement), 52*, 11-14.

Ogawa, A., & Macaluso, E. (2013). Audio–visual interactions for motion perception in depth modulate activity in visual area V3A. *NeuroImage, 71,* 158-167.

Pegna, A., Roberts, M., & Leek, C. (2014). The temporal dynamics of 3D object recognition for mono- and stereo visual displays: An ERP study. *Journal of Vision, 14*, 1294.

Peirce, J.W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Proctor, B. J., & Meyer, G. F. (2011). Electrophysiological correlates of facial configuration and audio–visual congruency: evidence that face processing is a visual rather than a multisensory task. *Experimental Brain Research, 213*(2-3), 203-211.

Regan, D., Erkelens, C.J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology and Vision Science, 27*, 584–597.

Regan, D., & Gray, R. (2000). Visually guided collision avoidance and collision achievement. *Trends in Cognitive Sciences, 4*(3), 99-107.

Roach, N. W., Heron, J. & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B: Biological Sciences, 273*(1598), 2159-2168.

Schiff,W., Caviness, J. A., & Gibson, J. J. (1962). Persistent fear responses in rhesus monkeys to the optical stimulus of 'looming'. *Science, 136*, 982–983.

Schlicht, E. J., & Schrater, P. R. (2007). Effects of visual uncertainty on grasping movements. Experimental Brain Research, 182, 47–57.

Sheedy, J. E., Bailey, I. L., Buri, M., & Bass, E. (1986). Binocular vs. monocular task performance. American Journal of Optometry and Physiological Optics, 63, 839 – 846.

Soto-Faraco, S., Kingstone, A. & Spence, C. (2004). Cross-modal dynamic capture: Congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 330-345.

Stekelenburg, J. J., & Vroomen, J. (2009). Neural correlates of audio-visual motion capture. *Experimental Brain Research, 198*(2-3), 383-390.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. London: The MIT Press.

Teder-Sälejärvi, W.A., McDonald, J.J., DiRusso, F., & Hillyard, S.A.(2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14*(1), 106–114.

Teder-Sälejärvi, W. A., Russo, F. D., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience, 17(9)*, 1396-1409.

Tresilian, J. R., Mon-Williams, M., & Kelly, B. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings of the Royal Society of London, 266*, 39-44.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1996). Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. Journal of Neurophysiology, 76, 2872– 2885.

Tyler, C. W. (1991). Cyclopean vision. In D. Regan (Ed.), Vision and visual dysfunction: Vol. 9. Binocular vision (pp. 38 –74). London, United Kingdom: MacMillan

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190-123.

von Saldern, S., & Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *The Journal of Neuroscience, 33*, 8841-8849.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *JOSA A, 4*(10), 2006-2021.

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience, 30*, 2662-2675.

Wheatstone, C. (1838). Contributions to the physiology of vision.—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. Philosophical Transactions of the Royal Society of London, 128, 371–394.

Yan, J. J., Lorv, B., Li, H., & Sun, H. J. (2011). Visual processing of the impending collision of a looming object: Time to collision revisited. *Journal of Vision, 11*(12), 7.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error-detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931-959

Zimmer, U., Itthipanyanan, S., & Woldorff, M. G. (2010). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *European Journal of Neuroscience, 31*(10), 1744-1754.