

# Capture of Auditory Motion by Vision Is Represented by an Activation Shift from Auditory to Visual Motion Cortex

Arjen Alink,<sup>1,2</sup> Wolf Singer,<sup>1,2</sup> and Lars Muckli<sup>3</sup>

<sup>1</sup>Max Planck Institute for Brain Research, D-60528 Frankfurt am Main, Germany, <sup>2</sup>Brain Imaging Center Frankfurt, D-60528 Frankfurt am Main, Germany, and <sup>3</sup>Department of Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom

The brain is capable of integrating motion information arising from visual and auditory input. Such integration between sensory modalities can aid one another and helps to stabilize the motion percept. However, if motion information differs between sensory modalities, it can also result in an illusory auditory motion percept. This phenomenon is referred to as the cross-modal dynamic capture (CDC) illusion. We used functional magnetic resonance imaging to investigate whether early visual and auditory motion areas are involved in the generation of this illusion. Among the trials containing conflicting audiovisual motion, we compared the trials in which CDC occurred to those in which it did not and used a region of interest approach to see whether the auditory motion complex (AMC) and the visual motion area hMT/V5+ were affected by this illusion. Our results show that CDC reduces activation in bilateral auditory motion areas while increasing activity in the bilateral hMT/V5+. Interestingly, our data show that the CDC illusion is preceded by an enhanced activation that is most dominantly present in the ventral intraparietal sulcus. Moreover, we assessed the effect of motion coherency, which was found to enhance activation in bilateral hMT/V5+ as well as in an area adjacent to the right AMC. Together, our results show that audiovisual integration occurs in early motion areas. Furthermore, it seems that the cognitive state of subjects before stimulus onset plays an important role in the generation of multisensory illusions.

**Key words:** audiovisual; fMRI; multisensory; cross-modal; motion; integration; MT

## Introduction

Both vision and audition allow us to extract important information about the velocity and direction of moving objects. Although the physical signals entering the two sensory systems are quite different, the brain is able to integrate unimodal information into a multimodal representation of motion. As shown by Meyer et al. (2005), audiovisual integration can increase the detection rate of motion when motion is colocalized across senses. This brings clear benefits in natural surroundings, where moving sounds and moving visual stimuli often belong to the same object. However, when visual and auditory stimuli move in opposite directions, audiovisual integration can lead to erroneous motion perception. This was demonstrated in an experiment by Soto-Faraco et al. (2002). They showed that when auditory apparent motion conflicts with visual motion, the perceived direction of the auditory apparent motion stimulus is inverted and made consistent to the visual, a phenomena that is referred to as cross-modal dynamic capture (CDC). The main aim of the functional magnetic reso-

nance imaging (fMRI) study at hand is to examine the neuronal basis of this illusion.

If vision alters auditory motion perception as demonstrated by cross-modal dynamic capture, at what stage of motion processing then does this interaction take place? Some claim that audiovisual motion interactions can be explained by decisional processes that occur after visual and auditory motion have been processed independently (decisional hypothesis) (Wuerger et al., 2003; Alais and Burr, 2004), whereas others claim that interactions take place during early visual and auditory motion processing (perceptual hypothesis) (Kitagawa and Ichihara, 2002; Soto-Faraco et al., 2005). Psychophysical studies have found support for both the decisional and the perceptual hypothesis (Kitagawa and Ichihara, 2002; Alais and Burr, 2004; Soto-Faraco et al., 2006; Sanabria et al., 2007). To further resolve this issue, we tested the decisional hypothesis by investigating whether early motion specialized areas are affected by cross-modal interactions.

So far, most fMRI studies on audiovisual motion integration have focused on identifying areas that are involved in both visual and auditory motion processing. Such areas were localized in frontal and parietal cortices by Lewis et al. (2000) and Bremmer et al. (2001). More relevant with respect to the integration of motion across modalities is a recent fMRI study by Baumann and Greenlee (2007) that examined the effects of audiovisual motion coherency. The present investigation goes beyond this study by assessing the occurrence of cross-modal dynamic capture during trials in which auditory and visual motion are conflicting. This allowed us to identify the neural mechanisms of cross-modal

Received June 30, 2007; revised Nov. 1, 2007; accepted Dec. 10, 2007.

This work was supported by the Federal Ministry of Education and Research (BMBF 01 GO 0508). We are very grateful to Tim Wallenhorst for his help with our data acquisition and to Michael Wibrall for helpful advice for the data analysis. We thank Fraser W. Smith and Scott Fairhall for helpful comments on this manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Arjen Alink, Department of Neurophysiology, Max Planck Institute for Brain Research, Deutschordenstrasse 46, D-60528 Frankfurt am Main, Germany. E-mail: alink@mpih-frankfurt.mpg.de.

DOI:10.1523/JNEUROSCI.2980-07.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/282690-08\$15.00/0

dynamic capture by comparing blood oxygenation level-dependent (BOLD) responses for CDC trials to nonillusory trials. To test the decisional hypothesis, we investigated whether this kind of integration takes place in specialized visual and/or auditory motion areas. To this end, we functionally localized the visual motion complex (hMT/V5+) and areas sensitive to auditory motion. Using the same region of interest (ROI)-based approach, we also tested whether audiovisual motion coherency influences modality-specific motion areas.

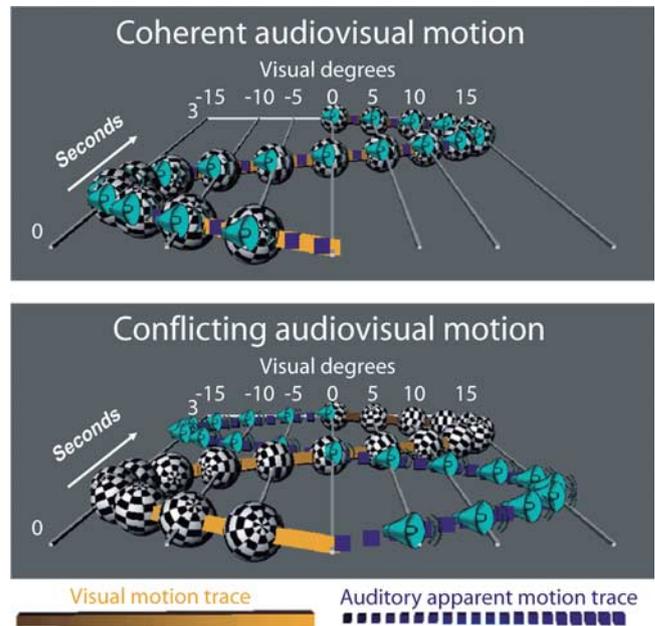
## Materials and Methods

### Subjects

Ten healthy volunteers participated in the main fMRI study (age range, 23–36 years; five females). Seven of these subjects (three females) participated in the localizer experiment. All subjects had normal hearing and normal or corrected-to-normal vision. All subjects gave their informed consent after being introduced to the experimental procedure in accordance with the declaration of Helsinki.

### Stimuli and task

**Main experiment.** Visual stimuli were presented using an MR-compatible goggle system with two organic light-emitting diode displays (MR Vision 2000; Resonance Technology, Northridge, CA), and auditory stimulation was performed using an MR-compatible head phone system (Commander XG; Resonance Technology). The screen had a width of 30° and a height of 22.5°, and the luminance of the gray background was 24.0 cd/m<sup>2</sup>. During the audiovisual trials, subjects were exposed to a moving sphere (1.5° radius) with a black and white checkerboard texture (luminance black = 1.2 cd/m<sup>2</sup>; luminance white = 43.9 cd/m<sup>2</sup>). At the same time, they heard a stream of 20 bass drum sounds (67.5–79.5 dB) with a duration of 100 ms presented with a 50 ms interstimulus interval. During these 20 periods of auditory stimulation, the sphere pulsed (radius increase of 0.45°), which resulted in strong perceptual binding between visual and auditory stimuli. The sounds induced an apparent motion percept that was realized by a transformation of the sounds by a head-related transfer function (HRTF) created by the Massachusetts Institute of Technology Media Lab Machine Listening Group (Gardner and Martin, 1994) with a precision of 5° (implemented with Matlab). Both visual and auditory stimuli moved sinusoidally on the horizontal midline of the screen from the center to 15° eccentricity of both sides and back to the center within 3 s while subjects fixated at a white fixation cross 3.75° below the center of the screen. During coherent audiovisual trials, the auditory apparent motion had the same direction as the visual stimulus, whereas the direction was opposite across senses during conflicting trials (Fig. 1). The task during both audiovisual conditions was to report on the initial motion direction of the auditory stimulus. Subjects had to respond with their right hand before the stimuli disappeared. An index finger press indicated initial leftward motion, and a middle finger press indicated initial rightward motion. After the stimuli disappeared, the fixation cross turned green in the case of a correct response and red in the case of an incorrect response, no response, or multiple responses. The fixation cross remained in this color until the end of the trial which had a total length of 4 s. For the unimodal trials, either visual or auditory stimuli were presented; the unimodal stimuli were essentially the same as those presented in the audiovisual trials. During the unimodal auditory trials, subjects had to respond to the initial motion direction of the auditory stimulus in the same way as in the audiovisual trials. During the visual trials, however, subjects had to respond to the initial visual motion direction. Before the experiment began, each subject completed 10 practice trials for each condition outside the scanner. Within the scanner, subjects completed four runs that each contained 25 trials per condition. Within each run, trials were intermixed with 25 fixation periods (4 s duration each), which served to assess the baseline signal. Thus, during an fMRI session, each subject was presented with 100 trials per condition over all runs. We used a rapid event-related paradigm. To allow for correct deconvolution of the BOLD responses for each condition, the history of trials was balanced in each run to control for an equal occurrence of the preceding two trials. This “two-back” balancing was achieved by draw-



**Figure 1.** A graphical representation of the audiovisual stimuli used in the experiment. Stimulation consisted of continuous visual motion and auditory apparent motion. The spheres represent the location of the visual stimuli, and the speakers represent the perceived location of the auditory stimuli over time during a trial. Auditory stimuli were presented with a spatial resolution of 5°. Stimuli locations are depicted for the 20 intervals during which an auditory stimulus was presented for 100 ms (interstimulus interval was 50 ms). An example is shown of a coherent trial in which audiovisual stimuli move coherently (top) and of a conflicting trial (bottom) in which motion direction was opposite across senses. During the experiment, these trials were presented together with an equal amount of their mirrored equivalents.

ing randomly a start sequence of three trials (triplet) and subsequently drawing a fitting triplet in which the first two trials match the last two trials of the preceding triplet. The drawing procedure was repeated until a valid solution was found that used all triplets (sequences were automatically generated using Matlab). The first triplet in each run was returned to the bowl and was drawn again at a random later position within the run. This resulted in a two-back balanced sequence per run that contained 25 trials for all five conditions (including fixation) plus the initial triplet that was disregarded in the data analysis. Furthermore, we ensured that the frequency of the initial motion direction of the visual and auditory motion stimuli was identical within and between conditions.

**Localizer experiment.** For localization of the human visual motion complex (hMT/V5+), we used a standard block design mapping procedure that we have used and described previously (Muckli et al., 2002). In short, we used expanding low-contrast random-dot-flowfield patterns covering a visual field of ~20° × 30° visual angle. Moving flowfields of random dot patterns (RDPs) were compared with static random dot patterns. Visual projection during the localizer experiment: the visual stimuli were back-projected onto a frosted screen attached to the end of the head coil, which subjects could see through a mirror mounted to the inside of the head coil. For the localization of motion-sensitive auditory cortex, the same HRTF function as in the main experiment was used to create a sound that appeared to rotate around the subjects head. For this, we used the first 18 s of guitar music by Jeff Wahl (title: Groove) because it contained only small amplitude changes over time and we considered it a relaxing and enjoyable sound for the subjects to focus on during the measurement. The sound rotated on the horizontal plane (0 azimuth) with a speed of 5° per second and was presented using the same headphones as in the main experiment. Using the left and right channel of this rotating stimulus, we created two stationary control stimuli. One of the stationary control stimuli consisted of the binaurally presented left channel and the other of the binaurally presented right channel of the rotating sound stimulus. Both of these stimuli gave rise to the perception of amplitude changes over time but did not induce any motion percept. For the

analysis, we pooled the two stationary sounds into one static control condition that contained the same stimulation across ears over all trials as the rotating stimulus. In total, we presented the subjects with 30 samples of the moving sound and 30 samples of the static control sounds (15 of each type). All samples had a duration of 18 s. The order of the trials was randomized, and sound samples were separated from one another by an 18 s fixation period. During the entire experiment, subjects were instructed to fixate at a central white fixation cross on a black screen.

### fMRI procedure

**Main experiment.** Functional and anatomical MRI data were acquired with a 3T-MRI system (Siemens Allegra; Siemens, Erlangen, Germany) using a four-channel head coil. For each subject, we obtained 516 volumes containing 20 slices covering the entire brain during each of the four functional scans using a gradient echo-echo planar imaging (EPI) sequence [repetition time (TR), 1000 ms; echo time (TE), 25 ms; flip angle, 70°; voxel size,  $3.4 \times 3.4 \times 5.0$  mm; field of view (FOV), 220 mm; gap thickness, 0.7 mm]. We corrected for spatial distortions in the EPI images using a point spread function (PSF) (Zaitsev et al., 2004). We also obtained a T1-weighted anatomical scan for each of the subjects using a Siemens magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence ( $1 \times 1 \times 2$  mm).

**Localizer experiment.** For this experiment, we used the standard Siemens 3T head coil. Functional data for visual motion mapping was acquired in one functional scan containing 216 volumes. For the auditory motion mapping, we measured three times 720 volumes in three scans. The same EPI sequence was used for visual and auditory motion mapping (TR, 1000 ms; TE, 30 ms; flip angle, 77°; voxel size,  $3.4 \times 3.4 \times 3.5$  mm; FOV, 220 mm; gap thickness, 0.35 mm), and PSF was applied to correct for spatial distortions. Slices were orientated parallel to the planum temporale covering the lower part of the parietal, the lower part of the frontal, the upper part of the temporal, and the entire occipital lobe. We also obtained a T1-weighted anatomical scan for each of the subjects using a Siemens MPRAGE sequence ( $1 \times 1 \times 1$  mm).

### Data analysis

**Main experiment.** Cross-modal dynamic capture trials were defined as conflicting trials in which the opposite of true motion direction was reported. All other trial types (coherent, visual, and auditory trials) with false responses were excluded from further analysis. Furthermore, the first three trials in each scan were excluded to preclude T1 saturation effects and to ensure a balanced two-back history across conditions. fMRI data were analyzed using the BrainVoyager QX software package (Brain Innovation, Maastricht, The Netherlands). Data were preprocessed using the default settings of BrainVoyager QX. After alignment with the anatomical reference scan, we spatially smoothed the functional data using a Gaussian kernel with a full-width at half-maximum of 8 mm. All individual datasets were transformed into Talairach space (Talairach and Tournoux, 1988). A group-based general linear model (GLM) was computed using a deconvolution design (Glover, 1999). *t*-value maps were computed for the following contrasts: coherent > conflicting, conflicting > coherent, CDC > conflicting, and conflicting > CDC. Contrasts were tested for significance over the expected interval of the BOLD response, which was 2–7 s after trial onset. A second contrast that tested whether the BOLD peak was above baseline was used in conjunction with the contrasts above. We used a threshold of  $t > 3.3$  ( $p < 0.001$ ) together with a cluster threshold that corrected for multiple comparisons ( $p$  corrected < 0.05). Cluster thresholds were computed for each contrast using the method introduced by Forman et al. (1995) and implemented in BrainVoyager QX by Fabrizio Esposito and Rainer Goebel (University of Maastricht, Maastricht, The Netherlands).

**Localizer experiment and ROI analysis.** All data were preprocessed and normalized as in the main experiment, although the data were not spatially smoothed. A group-based GLM was calculated for auditory and visual motion mapping using a single-factor design that used regression to the hemodynamic response function (Boynton et al., 1996). *t*-value maps were computed for the contrast moving auditory stimulus > stationary auditory stimulus for localization of auditory motion complex (AMC) and flowfield RDP > static RDP for the localization of hMT/

V5+. ROIs for hMT/V5+ were identified individually for the seven subjects that participated in the localizer experiment. These consisted of the  $1000 \pm 44$  mm<sup>3</sup> that was most significantly activated for the flowfield RDP > static RDP contrast near the posterior part of the inferior temporal sulcus. For the three subjects that could not participate in the localizer experiment, we defined hMT/V5+ ROIs as the  $1000 \pm 44$  mm<sup>3</sup> close to the posterior part of the inferior temporal sulcus that was most significantly activated by the visual motion condition in the main experiment. ROIs for AMC were defined on a group level of all seven subjects in the localizer experiment. We preferred group level ROIs to individual ROIs because AMC mapping was not robust enough at an individual level but was sound at a group level. Note that although this ROI was defined based on the data of seven subjects, we used it for the ROI analysis over all 10 subjects. Defining these regions of interest allowed us to generate event-related time courses for hMT/V5+ and AMC for the data in the main experiment using deconvolution. To test whether audiovisual integration took place in these motion areas, we tested the following contrasts for significance: coherent > conflicting, conflicting > coherent, cross-modal dynamic capture > conflicting, and conflicting > cross-modal dynamic capture. Results indicated that responses in hMT/V5+ for audiovisual stimulation were actually lower than those for visual stimulation alone. Furthermore, both motion areas turned out to respond to both unimodal auditory and unimodal visual stimulation. Therefore, we also tested the following contrasts for significance [visual > average (coherent + conflicting), visual > 0, and auditory > 0]. All contrasts were calculated over the onset and the peak of the BOLD responses, which lasted from 2 to 7 s after stimulus onset. This corresponded to a contrast over the data points 3 to 8 as shown in Figure 4 and supplemental Figure 1 (one data point/volume was recorded each second, the first being recorded at stimulus onset). To assess how consistent multisensory effects in the auditory and visual motion areas were over subjects, we plotted the effect sizes for the contrasts for CDC minus conflicting and coherent minus conflicting for each individual separately. To make sure that multisensory effects on a group level were not driven by a single subject, we tested whether there were extreme outliers in the individual data by testing for the presence of subjects with effect sizes that were beyond three times the interquartile range.

## Results

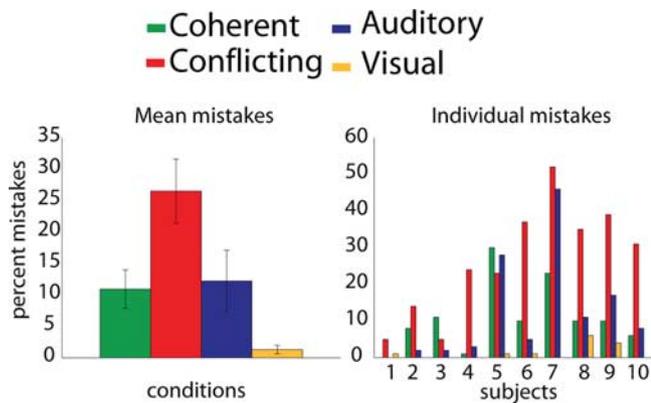
### Behavioral data

During the fMRI experiment, cross-modal dynamic capture occurred in 26.5% of the conflicting trials. This was significantly higher ( $p < 0.0005$ ) than the misclassifications in the congruent trials (10.9%) and the auditory trials (12.2%), which did not significantly differ from each other. The visual trials showed significantly lower misclassification than any other type of trial (1.3%;  $p < 0.0007$ ). Individual mistake rates varied widely over subjects. However, higher mistake rates for the conflicting condition turned out to be consistent with 8 of 10 subjects exhibiting this trend. See Figure 2 for more details.

### fMRI data

#### Localization of hMT/V5+

In our study, we focused on early motion-specific areas in the visual and auditory cortex. We therefore used standard mapping procedures to define motion-specific ROIs in visual and auditory motion cortex. For the visual modality, responses to moving low-contrast RDPs were compared with responses to static low-contrast RDPs in 7 of the 10 subjects. We localized the left and right human visual motion complex (hMT/V5+) for each of these subjects by selecting  $1000 \pm 44$  mm<sup>3</sup> of brain volume near the posterior part of the inferior temporal sulcus that was most motion sensitive. On the group average, the left and right hMT/V5+ was located at the following Talairach coordinates: left hMT/V5+,  $x = -44.1$  (SE, 1.5);  $y = -67.0$  (SE, 0.9);  $z = 0.7$  (SE, 1.6); right hMT/V5+,  $x = 45.3$  (SE, 2.1);  $y = -64.3$  (SE, 3.0);  $z =$



**Figure 2.** Left, The group average of the behavioral data showing the amount of misclassification during the different conditions (including SEs across subjects). During all types of trials in which auditory stimuli were presented, subjects reported the direction of auditory motion, whereas during the visual trials subjects responded to the direction of visual motion. Right, The misclassification rate for each single subject.

−0.3 (SE, 1.9). For the three subjects that could not participate in the localizer experiment, we used the activation for the visual motion condition in the main experiment to define a ROI for hMT/V5+. This area was always an island of activation in the vicinity of the posterior part of the inferior temporal sulcus. For this group, the ROI for the left hMT/V5+ was located on average at  $x = -49.3$  (SE, 3.3);  $y = -65.0$  (SE, 3.2);  $z = 2.3$  (SE, 1.5), and the location of the right hMT/V5+ ROI was  $x = 42.3$  (SE, 2.3);  $y = -62.0$  (SE, 0.7);  $z = 1.7$  (SE, 2.2).

#### Localization of AMC

For the auditory modality, responses to moving sounds that appeared to rotate around the subject's head were compared with static control conditions. We defined auditory motion-sensitive areas based on a group analysis. An area in the posterior part of the planum temporale in both hemispheres (Fig. 3) responded more strongly to the moving sound condition than to the static control condition. These areas are in agreement with the motion-sensitive areas found in a previous fMRI study by Baumgart et al. (1999). We refer to these areas as the AMC. One other area was found with the same preference for moving auditory stimuli, which was located in the left parietal cortex.

#### ROI-based hMT/V5+ analysis

The main question of our study was whether motion-sensitive areas, as defined by the ROI-mapping procedures, are affected by the cross-modal dynamic capture illusion. Therefore, we compared the ROI activity in trials in which the illusion of cross-modal dynamic capture was present with those (physically identical) trials in which no such illusion was perceived. Both the left and right visual motion complex (hMT/V5+) responded more strongly to conditions in which the illusion was subjectively present (left,  $p < 0.0004$ ; right,  $p < 0.04$ ). In other words, for those conditions in which the visual stimulus dominated the auditory percept, bilateral hMT/V5+ showed higher activity. In individual subjects, responses were consistently higher for CDC trials in left hMT/V5+, whereas this difference was less consistent in the right hMT/V5+ (Fig. 4b). What is the response of hMT/V5+ if the coherency is already given in the physical stimulus? Left and right hMT/V5+ both show stronger responses to coherent audiovisual stimulation compared with conflicting stimulation (left hMT/V5+,  $p < 0.00003$ ; right hMT/V5+,  $p < 0.002$ ). This effect of motion coherency was highly consistent across sub-

jects with all subjects having a higher BOLD response for the coherent condition in left hMT/V5+ and 9 of 10 in the right hMT/V5+. This is a strong indication that this visual motion area responds indirectly to auditory stimulation. Testing for outliers regarding the individual CDC and coherency effect sizes showed that there were no extreme outliers that drive the observed effects on a group level. Does hMT/V5+ also respond directly to auditory stimulation? We tested the response to purely auditory stimulation for statistical significance and found that left and right hMT/V5+ show such a response (left hMT/V5+,  $p < 0.00001$ ; right hMT/V5+,  $p < 0.00001$ ). The strongest response is, however, for pure visual stimulation (visual > audiovisual stimulation: left hMT/V5+,  $p < 0.00001$ ; right hMT/V5+,  $p < 0.00001$ ). Together, there is a clear influence of auditory stimulation on visual area hMT/V5+. Visual-induced activity decreases if an auditory stimulus is present and even more if this stimulus is conflicting. However, for the conflicting trials that are perceived as coherent (CDC trials), the decrease appears to be absent.

#### ROI-based AMC analysis

Like hMT/V5+, also AMC was affected by cross-modal dynamic capture. In contrast to the left hMT/V5+, both the left and the right AMC showed a decrease in activation when cross-modal dynamic capture took place (left AMC,  $p < 0.005$ ; right AMC,  $p < 0.006$ ). Thus, when the visual motion percept dominated over the auditory motion percept, there was a bilateral reduction in activation in auditory motion cortices. In the individual data, this effect was expressed in the majority of the subjects in the right AMC while being less consistent within the left AMC. Does AMC, like hMT/V5+, respond differently when physical stimuli move coherently compared with when they move in opposite directions? This did not turn out to be the case. As for hMT/V5+, we tested whether there were outliers driving the observed cross-modal dynamic capture effect. This was not the case for left and right AMC. Our data show that inside the AMC, responses to coherent audiovisual motion do not differ from responses to conflicting audiovisual motion. As for hMT/V5+, we tested whether there was a different response in AMC to audiovisual motion compared with the preferred unimodal motion stimuli, which is auditory motion for AMC. We did not find evidence for such a difference. Hence, in AMC responses were close to identical for auditory motion, coherent audiovisual motion, and conflicting audiovisual motion. Is AMC also activated by movement in the nonpreferred modality? Surprisingly, both left and right AMC responded robustly to the unimodal visual motion stimulus (left AMC,  $p < 0.00001$ ; right AMC,  $p < 0.00001$ ).

Together, the ROI-based results indicate that there is a clear shift in activation during cross-modal dynamic capture from auditory motion areas to visual motion areas. Furthermore, it indicates that visual motion areas prefer audiovisual coherent motion over audiovisual conflicting motion and that both visual and auditory motion areas respond to visual as well as auditory motion.

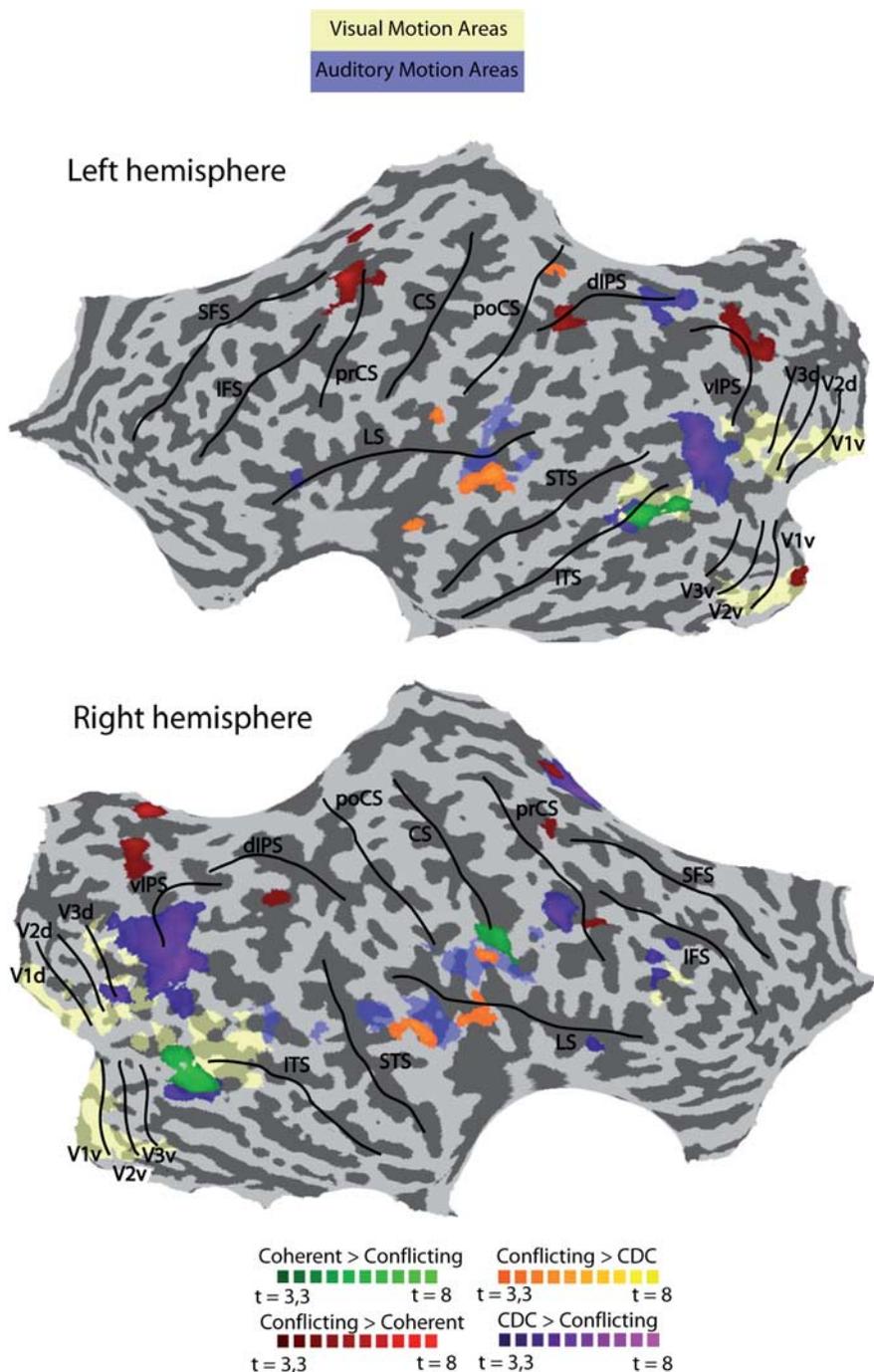
#### Full-brain analysis

Changing the focus of the study from an ROI-based approach to a whole-brain analysis allows for the identification of other brain areas that might also be tightly linked to the perceptual illusion of cross-modal dynamic capture. Furthermore, it can provide a more complete picture of brain regions that discriminate between coherent and conflicting audiovisual motion. First of all, the results of the ROI analysis were confirmed by the full-brain analysis: coherency preference was found in bilateral hMT/V5+

and consistent cross-modal capture effects were identified in bilateral hMT/V5+ and bilateral AMC. Although the full-brain analysis confirms that there is no coherency preference within AMC, it revealed that there is an area directly adjacent to the right motion-sensitive auditory cortex, located in the ventral part of the central sulcus, that shows a higher response when motion is coherent across senses. In addition to confirming cross-modal effects in hMT/V5+ and AMC, the full-brain analysis showed that conflicting motion led to higher responses in a large frontoparietal network comprising the bilateral precentral sulcus, the bilateral intraparietal sulcus, and right supplementary motor cortex. This effect was more pronounced in the left hemisphere, corresponding to the hand with which responses were made. To our surprise, we also found that conflicting motion also causes higher activation in primary visual cortex and the medial part of the thalamus. These effects were restricted to the left hemisphere. Our full-brain analysis also revealed areas other than bilateral AMC and bilateral hMT/V5+, in which activation was modulated by cross-modal dynamic capture. First of all, cross-modal dynamic capture was found to more strongly activate a large area within the left and right ventral intraparietal sulcus. Furthermore, cross-modal dynamic capture elevated responses within the left and the right insula, the left intraparietal sulcus, the right precentral sulcus, the right inferior frontal sulcus, and the right supplementary motor cortex. Higher activation for conflicting motion compared with cross-modal dynamic capture, in addition to being present in AMC, was also observed in the left superior temporal gyrus and the left precentral gyrus. The results of the full-brain analysis are summarized in Table 1.

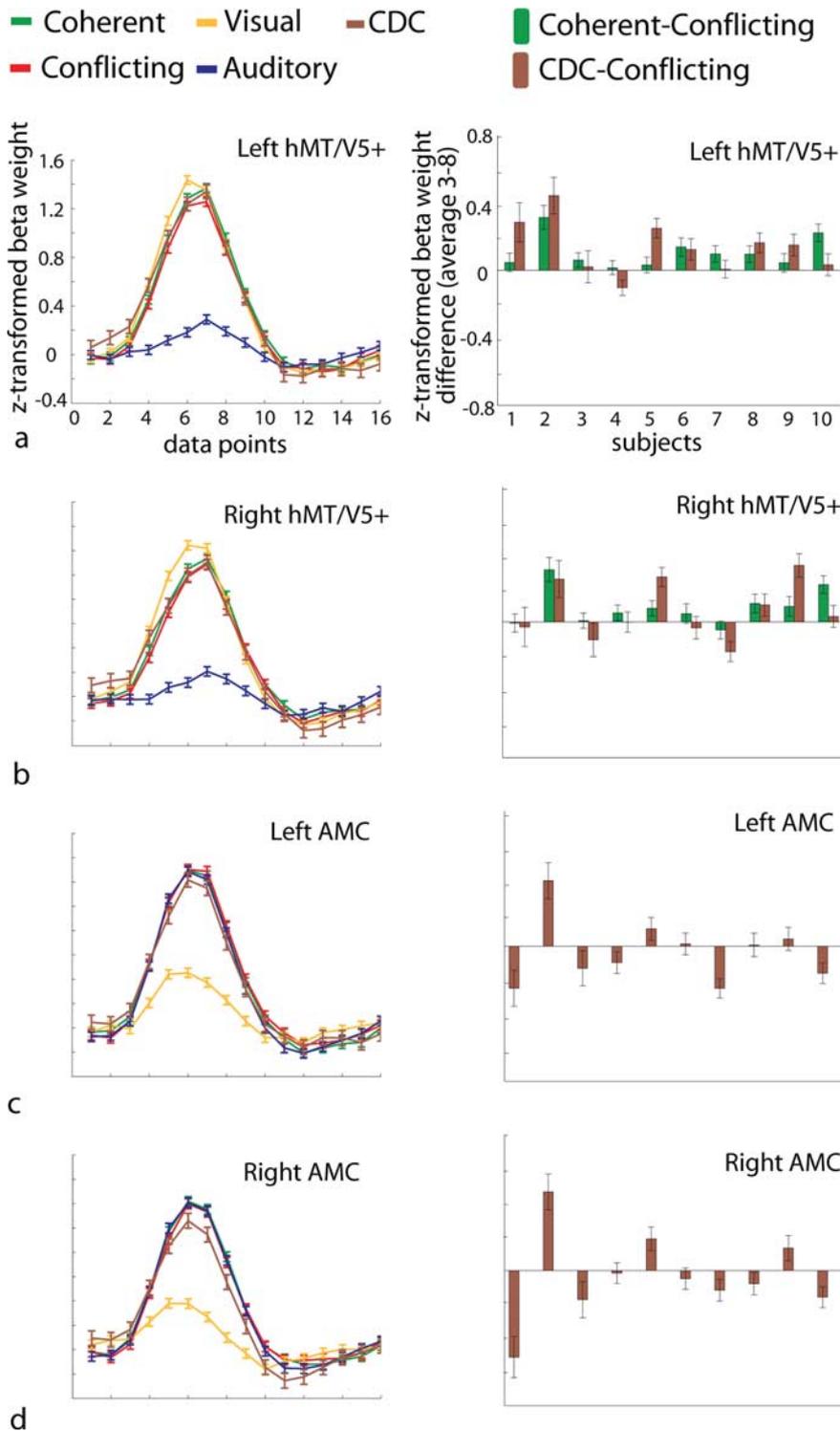
**Post hoc balanced history analysis**

In Figure 4, one can see that the BOLD responses of hMT/V5+ and AMC for the illusionary (CDC) condition is already higher than the other conditions for the first two data points after stimulus onset. This is earlier than can be explained by a typical stimulus-driven hemodynamic response. Because it is possible that this early offset carries over to the later stimulus-driven BOLD response, over which statistical tests were performed, it is important to show that it reflects neuronal processing that is relevant to audiovisual integration rather than being an artifact. To address this issue, we first examined to what extent this offset was present in all the other areas defined by the full-brain analysis (see supplemental Fig. 2, available at www.jneurosci.org as supplemental material). If the offset would be present equally in all these



**Figure 3.** Significantly activated regions displayed on a map representing the flattened cortex of the left and right hemisphere of subject IEN24. Areas are displayed that show significant effects ( $p < 0.05$  corrected) for the following contrasts: coherent > conflicting (green), conflicting > coherent (red), conflicting > cross-modal dynamic capture (orange), and cross-modal dynamic capture > conflicting (purple). The flat map further includes areas that show significant motion sensitivity ( $p < 0.005$  corrected) for visual motion and auditory motion using the contrasts flowfield RDP > static RDP for visual motion (yellow) and rotating sound > static sound for auditory motion (blue). CS, Central sulcus; dIPS, dorsal intraparietal sulcus; IFS, inferior frontal sulcus; ITS, inferior temporal sulcus; LS, lateral sulcus; poCS, postcentral sulcus; prCS, precentral sulcus; SFS, superior frontal sulcus; STS, superior temporal sulcus; vIPS, ventral intraparietal sulcus; V1d–V3d, dorsal visual areas 1–3; V1v–V3v, ventral visual areas 1–3.

areas, this would be an indication that it reflects a process that is highly unspecific and therefore is likely to be an artifact. It turned out that this was not the case. The offset was most pronounced in occipital and parietal areas, being most dominantly expressed in the bilateral ventral intraparietal sulcus. The offset was expressed only to a lesser extent in temporal regions while being fully absent



**Figure 4.** Left, The grand mean event-related responses generated using deconvolution for the ROIs hMT/V5+ and the AMC as defined in the localizer experiment. The ROI for the hMT/V5+ was defined on an individual level and the AMC on a group level. Event-related responses are shown for coherent audiovisual trials (green), conflicting audiovisual trials with correct responses (red), and conflicting audiovisual trials with wrong responses, i.e., inducing the illusion of cross-modal dynamic capture (brown), auditory (blue), and visual (yellow) trials. Right, Individual differences of the peak of the BOLD response (mean of data points 3–8) between coherent and conflicting trials (green) and the difference between cross-modal dynamic capture and conflicting trials (brown). Data points were recorded each second, the first starting at stimulus onset.

in all frontal areas except for the bilateral precentral sulcus. Thus, the higher BOLD signal at stimulus onset of CDC trials appears to be regionally specific rather than omnipresent.

One explanation for the observed offset could be an unbal-

anced history for the CDC trials. Hence, because the occurrence of the CDC trial occurred unpredictably during conflicting trials, we were not able to balance the history two-back, as we did for all other trial types, neither for the CDC condition nor for the conflicting condition. This can be a problem because in our rapid event-related design an unbalanced history can result into a spilling over of high BOLD amplitudes for one condition into another condition (Glover, 1999). For example, it could be that the offset for the CDC trials in parieto-occipital cortex is the result of a higher frequency of visual trials, which yield higher responses in occipital areas, in the recent past of the CDC condition as compared with that of conflicting trials in which no illusion takes place. To test whether an unbalanced history could explain our results, we performed a *post hoc* analysis for the areas hMT/V5+ and AMC in which we removed 115 of the 265 CDC trials and 50 of the 735 conflicting trials for the conflicting condition. These trials were chosen such that the analysis over the remaining trials was perfectly two-back balanced. To make sure that most trials were included in the analysis, we performed this analysis 20 times, selecting the trials to be removed in a randomized manner from all runs in all subjects. The average result of these 20 analyses is plotted in the supplemental Figure 1 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). In this figure, one can see that the higher BOLD signal at stimulus onset persisted both in hMT/V5+ and AMC after the balancing of the history. Therefore, this offset cannot be attributed to an imbalance of condition history. To exclude the possibility that an imbalance in history drives our multisensory effects in hMT/V5+ and AMC, we tested whether the effects observed in the original analysis were still present in the averaged balanced analysis (one-sided test). This turned out to be the case. Our balanced analysis confirmed all multisensory effects in the original analysis (coherency effect left hMT/V5+,  $p < 0.004$ ; coherency effect right hMT/V5+,  $p < 0.05$ ; CDC effect left hMT/V5+,  $p < 0.003$ ; CDC effect right hMT/V5+,  $p < 0.04$ ; CDC effect left AMC,  $p < 0.01$ ; CDC effect right AMC,  $p < 0.002$ ).

From the fact that the higher signal for CDC trials at stimulus onset is regionally specific and persists even when the history for all trial types is balanced, we conclude that this offset reflects a neuronal event dominantly present in occipital and parietal cortex before stimulus onset that is involved in generating the cross-modal dynamic capture illusion.

**Table 1. Summary of full-brain analysis**

Region	Hemisphere	Talairach coordinates			mm <sup>3</sup>
		x	y	z	
<b>Coherent &gt; conflicting</b>					
Lateral occipital cortex (LOC)	L	-45	-67	-4	1580
LOC	R	41	-72	-5	1375
Ventral central sulcus	R	59	-13	19	576
<b>Conflicting &gt; coherent</b>					
Medial parietal cortex (mPC)	L	-7	-70	40	1376
mPC	R	7	-60	43	2354
Dorsal intraparietal sulcus (dIPS)	L	-33	-43	36	624
dIPS	R	29	-52	41	538
Visual cortex	L	-9	-84	-4	391
Medial thalamus	L	-3	-7	-16	454
Precentral sulcus (prCS)	L	-29	-8	51	1799
Dorsal prCS	L	-9	-4	66	301
prCS	R	32	-9	61	161
Ventral prCS	R	37	6	31	164
Supplementary motor cortex (SMA)	R	2	7	49	183
<b>CDC &gt; conflicting</b>					
LOC	L	-48	-63	-3	1500
LOC	R	36	-66	-8	1861
Ventral intraparietal sulcus	L	-33	-80	17	4659
Ventral intraparietal sulcus	R	30	-74	33	7785
prCS	R	46	2	31	2450
Inferior frontal sulcus	R	44	28	25	974
SMA	R	4	11	56	2739
Insula (IS)	L	-28	21	10	198
IS	R	35	18	8	238
dIPS	L	-18	-67	55	1771
<b>Conflicting &gt; CDC</b>					
Planum temporale (PT)	L	-50	-27	11	2072
PT	R	59	-21	10	1959
Superior temporal gyrus	L	-44	-11	-7	290
Precentral gyrus	L	-26	-45	52	365

Listed are the significantly activated regions ( $p < 0.05$  corrected) for the following contrasts: coherent > conflicting, conflicting > coherent, conflicting > cross-modal dynamic capture, and cross-modal dynamic capture > conflicting. L, Left; R, right.

## Discussion

In this fMRI study, we investigated the neural correlates of cross-modal dynamic capture, which is the illusory percept of an auditory stimulus moving in the same direction as a simultaneously presented visual stimulus, while the auditory stimulus is physically moving in the opposite direction. We explored the underlying neural mechanisms of this illusion by comparing BOLD responses to two physically identical audiovisual trials: trials in which auditory motion was correctly perceived as moving in the opposite direction as the visual stimulus and cross-modal dynamic capture trials in which subjects falsely perceived the auditory stimulus as moving coherently with the visual stimulus. We wanted to discover whether cross-modal dynamic capture has an influence on early motion processing areas. To this end, we functionally defined ROIs for the visual motion area hMT/V5+ and the AMC. Results show a robust influence of cross-modal dynamic capture on both AMC and hMT/V5+: it reduced activation in bilateral AMC, whereas it resulted in an increase in activation in the bilateral hMT/V5+. Our full-brain analysis revealed that in addition to hMT/V5+, also cortex in the bilateral ventral intraparietal sulcus and in the left dorsal intraparietal sulcus exhibit an elevated response during cross-modal dynamic capture. Thus, it appears that when vision dominates the motion percept, hMT/V5+ and other visual areas are more strongly activated at the cost of activation levels in auditory motion cortex. We interpret this activation shift as resulting from a competition between early visual and auditory areas for the final

motion percept. In this light, cross-modal dynamic capture can be regarded as vision winning the competition between the two senses. These findings fit well with a recent fMRI-EEG experiment by Bonath et al. (2007). They assessed the neural correlates of the audiovisual ventriloquism illusion and, like in our study, observed a decrease in BOLD amplitude for the illusory trials compared with the correctly perceived trials in the posterior planum temporale. Furthermore, our observation that multisensory effects take place both in unisensory and multisensory brain regions appears to be in good agreement with the findings by Noeselt et al. (2007) on audiovisual temporal correspondence.

Interestingly, we observed an elevated BOLD signal already at time 0 for illusory trials in both AMC and hMT/V5+. Our *post hoc* analysis showed that this offset does not reflect inadequate deconvolution as a result of an unbalanced trial history. Therefore, we interpret this offset as a neuronal event taking place before stimulus onset, which is involved in the generation of the cross-modal dynamic capture illusion. We observed that this pre-stimulus activation is most dominantly present in the ventral intraparietal sulcus, which is an area known to be involved in visual, auditory, and tactile motion processing (Grefkes and Fink, 2005) and therefore is thought to play an important role in the integration of motion across modalities (Lewis et al., 2000). Thus, the observed enhanced activation before and during cross-modal dynamic capture trials fits well with the current knowledge about the ventral intraparietal sulcus. But why should this enhancement take place before the onset of the audiovisual stimulus? This can be accounted for by the fact that subjects had to report the auditory motion direction, whereas the visual modality had to be ignored, which by default influences the spatial percept of auditory stimuli (Alais and Burr, 2004). Therefore, they had to constantly suppress the natural influence of vision on audition to perform the task correctly. In this light, one can see the enhanced activation in the ventral intraparietal sulcus before stimulus onset as a breakdown of the suppression of audiovisual integration just before the cross-modal dynamic capture illusion trials. In short, we interpret the enhanced activation in the ventral intraparietal sulcus before stimulus onset as a spontaneous return to the natural state of the brain in which vision “aids” auditory motion processing.

Recently, the cortical responses to coherent and incoherent audiovisual stimulation were investigated by Baumann and Greenlee (2007). We have followed the same line of comparison by selecting those trials that were perceived as conflicting (no CDC) and compared them to coherent trials. As in the analysis for cross-modal dynamic capture effects, we first asked whether early motion-sensitive regions are affected by audiovisual motion coherency. Our ROI-based analysis showed that early visual motion areas were affected by motion coherency across senses; both in the left as well as in the right hMT/V5+, we observed a signal increase for coherent audiovisual motion. Within the AMC, we did not observe any effect of motion coherency, although the full-brain analysis does show that an area slightly superior to the right AMC shows an increased BOLD response when motion is coherent across senses. These results show that already in the early visual motion area hMT/V5+, motion information is integrated across senses. For the auditory modality, this integration appears to take place in areas beyond primary and motion-sensitive regions.

Furthermore, our full-brain analysis showed that when motion is conflicting, this activates several areas in frontal and parietal cortices more strongly compared with coherent motion. This effect was more pronounced in the left hemisphere, which was the

hemisphere directing the execution of the responses; subjects always responded with the right hand to indicate in which direction they thought the auditory stimulus was moving using the index and middle finger. Thus, it is likely that the effect in frontoparietal cortex is related to response selection. We interpret this effect as reflecting an increased demand on decisional processes involved in preparing the motor response when motion is conflicting across senses. Thus, subjects needed to apply more effort for choosing the correct response when a distracting motion stimulus in the visual modality was presented. The finding of the same effect in the left primary visual cortex and left thalamus indicates that these lateralized decisional processes involve some sort of visual gating mechanism. Similar audiovisual gating by the thalamus was also found also in a recent fMRI study by Baier et al. (2006).

Although we did not observe a coherency effect within AMC, our finding of an area that prefers coherent over conflicting motion just superior to the right AMC seems to fit reasonably well with the finding of Baumann and Greenlee (2007) regarding the increase in activation for coherent audiovisual motion in the supramarginal gyrus. However, Baumann and Greenlee did not observe a widespread increase of activation for conflicting motion in frontal and parietal cortices, as we did. This can be explained by differences in the paradigms used across studies. In the study of Baumann and Greenlee, subjects attended visual motion, whereas in our design subjects attended auditory motion. It has been shown that salient visual stimuli influence the perceived location of the auditory stimulus more strongly than vice versa (Alais and Burr, 2004). Therefore, it is not surprising that in the study by Baumann and Greenlee, behavioral results obtained during the fMRI experiment showed no effect of coherency on the performance in detecting visual motion during their fMRI experiment. In our experiment, however, behavioral results do show a clear effect of coherency on task difficulty. Therefore, it would be consistent with our interpretation that the absence of the frontoparietal effect in their study is related to the lower influence of auditory motion on the visual motion percept compared with the effect of visual motion on the auditory motion percept.

Our finding that early visual and auditory motion areas as well as frontal areas are affected by motion coherency and/or cross-modal dynamic capture suggests that both the perceptual and the decisional stage of motion processing are involved in the integration of motion across modalities. This appears to be in agreement with the recent finding by Sanabria et al. (2007) that the presence of a visual motion distractor shifted both the sensitivity ( $d'$ ) and the response criterion ( $c$ ) for the detection and classification of auditory apparent motion. However, the fact the cross-modal dynamic capture illusion is preceded by elevated activity in the ventral intraparietal sulcus supports more the perceptual explanation of multisensory integration, because this region has been associated more to sensory motion processing than decisional processes. However, it could be that more frontal areas, although showing a less elevated activation level, drive the effect in the ventral intraparietal sulcus. To resolve this issue, one would have to further investigate prestimulus brain activation using methods like MEG and EEG that have a finer temporal resolution than fMRI.

This study is the first to demonstrate the neuronal correlates of cross-modal dynamic capture. Our main finding is that this illusion is accompanied by a decrease in activation in the auditory motion complex, whereas hMT/V5+ increases its level of activation. We assume that these effects represent competition between senses for the final motion percept at an early level of motion

processing. Our data suggests that this competition is influenced by neuronal events in occipital and parietal cortex before stimulus onset. Thus, we would like to stress the point that cognitive states before stimulus onset could play a significant role in generating multisensory illusions.

## References

- Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.
- Baier B, Kleinschmidt A, Muller NG (2006) Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *J Neurosci* 26:12260–12265.
- Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. *Cereb Cortex* 17:1433–1443.
- Baumgart F, Gaschler-Markefski B, Woldorff MG, Heinze HJ, Scheich H (1999) A movement-sensitive area in auditory cortex. *Nature* 400:724–726.
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hilliard SA (2007) Neural basis of the ventriloquist illusion. *Curr Biol* 17:1697–1703.
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221.
- Bremmer F, Schlack A, Duhamel JR, Graf W, Fink GR (2001) Space coding in primate posterior parietal cortex. *NeuroImage* 14:S46–S51.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
- Gardner B, Martin K (1994) HRTF measurements of a KEMAR dummy-head microphone. Retrieved February 10, 2008 from <http://sound.media.mit.edu/KEMAR.html>.
- Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* 9:416–429.
- Grefkes C, Fink GR (2005) The functional organization of the intraparietal sulcus in humans and monkeys. *J Anat* 207:3–17.
- Kitagawa N, Ichihara S (2002) Hearing visual motion in depth. *Nature* 416:172–174.
- Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:873–888.
- Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localization. *Exp Brain Res* 166:538–547.
- Muckli L, Kriegeskorte N, Lanfermann H, Zanella FE, Singer W, Goebel R (2002) Apparent motion: event-related functional magnetic resonance imaging of perceptual switches and States. *J Neurosci* 22:RC219(1–5).
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431–11441.
- Sanabria D, Spence C, Soto-Faraco S (2007) Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: a signal detection study. *Cognition* 102:299–310.
- Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res Cogn Brain Res* 14:139–146.
- Soto-Faraco S, Spence C, Kingstone A (2005) Assessing automaticity in the audiovisual integration of motion. *Acta Psychol (Amst)* 118:71–92.
- Soto-Faraco S, Kingstone A, Spence C (2006) Integrating motion information across sensory modalities: the role of top-down factors. *Prog Brain Res* 155:273–286.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.
- Wuerger SM, Hofbauer M, Meyer GF (2003) The integration of auditory and visual motion signals at threshold. *Percept Psychophys* 65:1188–1196.
- Zaitsev M, Hennig J, Speck O (2004) Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magn Reson Med* 52:1156–1166.