

Natural, Metaphoric, and Linguistic Auditory Direction Signals Have Distinct Influences on Visual Motion Processing

Sepideh Sadaghiani, Joost X. Maier, and Uta Noppeney

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

To interact with our dynamic environment, the brain merges motion information from auditory and visual senses. However, not only “natural” auditory MOTION, but also “metaphoric” de/ascending PITCH and SPEECH (e.g., “left/right”), influence the visual motion percept. Here, we systematically investigate whether these three classes of direction signals influence visual motion perception through shared or distinct neural mechanisms. In a visual-selective attention paradigm, subjects discriminated the direction of visual motion at several levels of reliability, with an irrelevant auditory stimulus being congruent, absent, or incongruent. Although the natural, metaphoric, and linguistic auditory signals were equally long and adjusted to induce a comparable directional bias on the motion percept, they influenced visual motion processing at different levels of the cortical hierarchy. A significant audiovisual interaction was revealed for MOTION in left human motion complex (hMT+/V5+) and for SPEECH in right intraparietal sulcus. In fact, the audiovisual interaction gradually decreased in left hMT+/V5+ for MOTION > PITCH > SPEECH and in right intraparietal sulcus for SPEECH > PITCH > MOTION. In conclusion, natural motion signals are integrated in audiovisual motion areas, whereas the influence of culturally learnt signals emerges primarily in higher-level convergence regions.

Introduction

Integrating motion information across the senses enables us to interact effectively with our natural environment. For instance, in foggy weather, we may more easily detect a fleeing animal and discriminate the direction into which it is running by combining the degraded visual information and the sound of its footsteps. In everyday life, visual motion cues are more reliable and dominate the motion percept. Vision even reverses the perceived direction of a conflicting auditory motion stimulus, a phenomenon referred to as visual capture (Mateeff et al., 1985; Soto-Faraco et al., 2002; Soto-Faraco et al., 2004). However, if the visual stimulus is rendered unreliable, auditory motion can also bias the perception of visual motion direction (Meyer and Wuerger, 2001). Audiovisual (AV) motion interactions rely on an absolute spatiotemporal reference frame tuned to our natural environment. They are maximized when auditory and visual stimuli are temporally synchronous and spatially colocalized (Meyer et al., 2005). In addition to “natural” AV motion interactions, a recent study demonstrated that de/ascending pitch with no spatial motion information alters visual (V) motion perception as in a “metaphoric” auditory capture, i.e., gratings with ambiguous motion were more likely to be perceived as upward motion when accompanied by ascending pitch (Maeda et al., 2004). Even “linguistic” stimuli (e.g., spoken words “left vs right”) that entertain only

arbitrary relationships to physical dimensions through cultural learning were shown to bias the perceived motion direction (Maeda et al., 2004). Interestingly, whereas the effect of pitch was maximized when sounds were temporally overlapping with the visual stimulus, the influence of speech was maximal for words presented 400 ms after onset of the visual stimulus. These behavioral dissociations suggest that the influence of natural, metaphoric, and linguistic auditory direction signals on motion discrimination may arise at multiple processing stages ranging from perceptual to decisional. Similarly, at the neural level, they may emerge at different levels of the cortical hierarchy involved in AV motion processing. Using conjunction analyses, previous functional magnetic resonance imaging (fMRI) studies have associated integration of tactile, visual, and auditory motion with anterior intraparietal and ventral premotor cortices (Bremmer et al., 2001). However, conjunction analyses define a region as multisensory if it responds individually to both types of unimodal inputs. Growing evidence for early multisensory integration in “putatively unisensory” areas (van Atteveldt et al., 2004; Ghazanfar et al., 2005; Kayser et al., 2005; Ghazanfar and Shroeder, 2006; Bizley et al., 2007; Martuzzi et al., 2007) raises the question of whether additional “modulatory” integration processes (interactions), in which one unimodal input does not elicit a significant regional response in itself but rather modulates the response to the other modality could be found in “classical” visual [e.g., human motion complex (hMT+/V5+)] or auditory (e.g., planum temporale) motion areas.

The present study investigates the influence of natural (MOTION), metaphoric (PITCH), and linguistic (SPEECH) auditory direction signals on visual motion processing. In a visual-

Received Nov. 11, 2008; revised Jan. 12, 2009; accepted April 2, 2009.

This work was supported by the Deutsche Forschungsgemeinschaft and the Max Planck Society.

Correspondence should be addressed to Uta Noppeney, Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076 Tübingen, Germany. E-mail: uta.noppeney@tuebingen.mpg.de.

DOI:10.1523/JNEUROSCI.5437-08.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/296490-10\$15.00/0

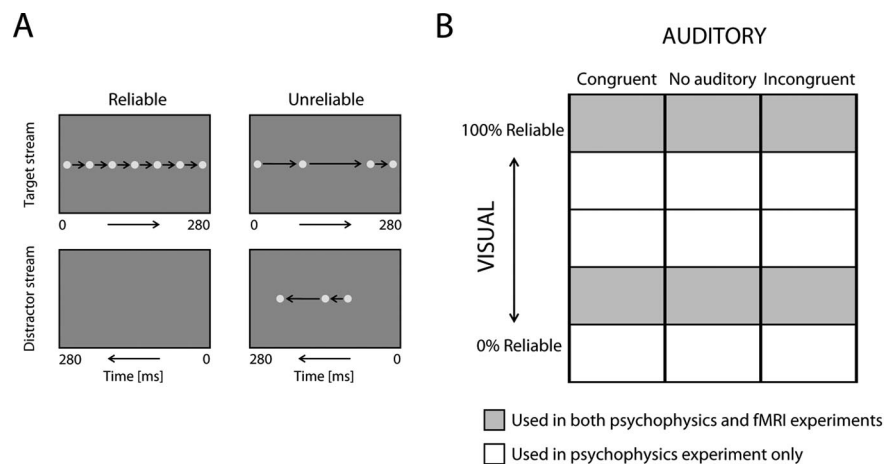


Figure 1. Visual stimuli and experimental design. **A**, Visual stimuli consisted of a single dot, presented sequentially along a target trajectory (target stream). Unreliability was introduced by adding a second dot moving along the same trajectory but in the opposite direction (distractor stream). For illustration purposes, only seven (instead of 14) dot positions are represented. For details, see Materials and Methods. **B**, In a visual-selective attention paradigm, subjects discriminated the direction of the visual motion. Each experiment (i.e., MOTION, PITCH, and SPEECH) conformed to a factorial design manipulating (1) reliability of the visual motion direction (5 or 2 levels) and (2) the auditory direction signal (3 levels: directionally congruent, absent, or incongruent to the visual signal).

selective attention paradigm, subjects discriminated the direction of visual motion at several levels of visual reliability, with an irrelevant auditory stimulus being congruent, absent, or incongruent. We then investigated whether the neural systems underlying audiovisual integration differ for natural, metaphoric, and linguistic contexts.

Materials and Methods

Subjects

After giving informed consent, 13 healthy German native speakers (seven females; median age, 23.5 years; one left-handed) participated in the psychophysical study, and 21 healthy right-handed German native speakers (six females; median age, 24 years) participated in the fMRI study. Nine of them participated in both studies. All subjects had normal or corrected-to-normal vision. The study was approved by the joint human research review committee of the Max Planck Society and the University of Tübingen.

Experimental design

In a visual-selective attention paradigm, subjects discriminated the direction of an apparent visual motion stimulus at several levels of reliability (or ambiguity) while ignoring a simultaneous auditory direction signal that could be congruent, absent, or incongruent. In consecutive experiments, audiovisual interaction and (in)congruency effects were investigated in three distinct contexts as defined by the auditory direction signal: (1) natural context (physical direction information), auditory MOTION signal; (2) metaphoric context (nonphysical, nonlinguistic direction information), de/ascending PITCH; and (3) linguistic context (nonphysical, linguistic direction information), SPEECH.

In each of the three experiments, the activation conditions conformed to a 5×3 (psychophysics) or 2×3 (fMRI experiment) factorial design manipulating (1) visual reliability (five or two levels) and (2) auditory direction signal (three levels). (1) The visual direction signal was presented at five or two levels of reliability ranging from intact to completely ambiguous. (2) The auditory direction signal was (a) indicative of leftward/downward direction, (b) absent, or (c) indicative of rightward/upward direction (compare with psychometric curves and estimation of directional bias). Alternatively, the levels of the auditory direction signal can be defined as (a) congruent, (b) absent/unimodal, or (c) incongruent to the direction of the visual motion (compare with analysis of accuracy, reaction time, and fMRI data).

In the following, the visual and auditory stimuli will be described in more detail.

Visual stimulus

To create an apparent visual motion signal, a single dot (1° visual angle diameter) was briefly (20 ms duration) presented sequentially at multiple locations along a horizontal or vertical trajectory (“target” stream). Ambiguous or unreliable motion direction signal was generated by presenting a second dot that followed the same trajectory but in the opposite direction (“distractor” stream) (Fig. 1A). This apparent visual motion stimulus is characterized by two features that are important for studying AV integration of motion information. First, the presentation of a spatially confined visual dot enables precise audiovisual colocalization, which is thought to be crucial for natural AV integration (Meyer et al., 2005). In contrast, the more frequently used random dot kinematogram does not provide audiovisual colocalization cues. Second, the presentation of a second interfering motion stream allows us to selectively manipulate motion information rather than low-level stimulus features (e.g., contrast or luminance of the dots) and imitates the effect of motion coherence in classical random dot

kinematograms. Thus, the apparent motion stimulus of the current study (1) provides the colocalization cues as given by natural moving physical objects and (2) enables selective manipulation of motion information similar to random dot kinematograms.

In each trial, the total number of flashes, i.e., dot presentations, was 14. The motion direction signal was parametrically rendered unreliable, by reducing the number of flashes in the target stream but increasing the number of flashes by an equal amount in the distractor stream. In the psychophysics study, five levels of reliability were introduced by assigning the flashes to the target versus distractor streams according to the following ratios: (1) 14:0 (completely reliable), (2) 10:4, (3) 9:5, (4) 8:6, and (5) 7:7 (completely ambiguous/unreliable). In the fMRI study, only levels 1 and 4 were used.

For both the horizontal and vertical trajectories, 14 equidistant (1.8° distance) potential presentation locations were used ranging from -11.7 to $+11.7^\circ$. In each trial, these 14 locations were randomly assigned to either the target or distractor streams. This ensured that, in each trial, a flash (dot) was presented exactly once at each prespecified location. In other words, to control the reliability and amount of visual input (luminance), we excluded the possibility that a target and distractor dot were simultaneously presented at the same location. However, although the potential locations were equidistant, obviously the actual flashes of a particular stream in a specific trial were not. Hence, to hold the speed ($89^\circ/\text{s}$) of the apparent motion constant, the timing of the flashes were adjusted to the appropriate spatial distance; 20 ms were dedicated to each distance between two possible locations along a stream regardless of whether or not a flash was actually presented at this location in a particular trial.

The duration of each apparent motion stimulus was 280 ms (i.e., 14×20 ms) and followed by the next trial after an additional intertrial interval of 1500 ms, yielding a stimulus onset asynchrony of 1780 ms.

In both the psychophysical and the fMRI studies, the dot was presented in light gray on a dark gray background at a luminance contrast of 0.5 (Weber contrast) as measured by a Minolta chroma meter (model CS-100). In the psychophysical study, the absolute luminance of the dot and the background were 19.9 and 10 cd/m^2 , respectively. Absolute measures of luminance cannot be accurately reported for the projection screen inside the MR-scanner because of a measurement distance of ~ 3 m (attributable to magnetic incompatibility of the measurement device).

Auditory stimulus

The three experiments used three distinct classes of auditory direction signals, as follows.

Auditory MOTION. Auditory MOTION is a white noise stimulus moving from left (-11.7°) to right ($+11.7^\circ$) or vice versa. In the MOTION experiment, both the auditory and visual motion signals followed a horizontal trajectory, because human auditory motion discrimination is much more precise for horizontal than vertical directions (Saber and Perrott, 1990). The auditory apparent motion stimulus was a white-noise stimulus of 20 ms duration that was played subsequently at each of the 14 possible locations at which the visual stimulus was presented. To allow for precise spatiotemporal colocalization of the auditory and visual motion signals and hence enable natural AV integration in the scanner environment, before the experiment, the white-noise stimulus was played from each of the 14 equidistant (1.8° distance) locations ranging from -11.7 to $+11.7^\circ$ and recorded individually for each subject using ear-canal microphones. With the help of this procedure, the auditory stimulus did not only provide binaural sound localization cues such as interaural time and intensity difference but also subjective monaural filtering cues of sound structure by the head, shoulders, and outer pinna (Pavani et al., 2002).

De/ascending PITCH. De/ascending PITCH is a linearly rising (from 200 Hz to 3 kHz) or falling (from 3 kHz to 200 Hz) pitch. In the PITCH experiment, the visual motion was vertical, because rising and falling pitch influences perception of vertical visual motion (Maeda et al., 2004).

SPEECH. SPEECH included the spoken German words “links” (left) and “rechts” (right) or “auf” (up) and “runter” (down). In the SPEECH experiment, one experimental session presented horizontal and the other one vertical visual motion together with the appropriate German words denoting left/right or up/down. This allowed us to compare formally whether the direction of the visual signal per se (i.e., vertical vs horizontal) influenced AV interactions, when the context (or auditory direction signal) was held constant.

The duration of all auditory stimuli was set to 280 ms to match the duration of the visual motion stream. The auditory direction signal was never rendered unreliable. Based on extensive psychophysical piloting (e.g., audiovisual stimulus duration or PITCH frequency), the auditory signals were adjusted to induce a comparable bias across the MOTION, PITCH, and SPEECH contexts.

Functional motion localizer

To localize the human visual motion complex (hMT+/V5+), a standard block-design MT+ (containing both areas MT and MST) localizer was performed (Krekelberg et al., 2005). In brief, in each of the 10 blocks, 10 s of expanding/contracting radial random dot motion (speed of $4^\circ/s$, reversal rate of 1 Hz, aperture of 20° visual angle) alternated with epochs of stationary random dots (24 s) on a dark screen. The dots were presented in maximal-luminance white and the background in minimal-luminance black, resulting in 100% luminance contrast. Subjects fixated a central fixation square that, at random intervals, changed its luminance (dark vs light gray). To maintain attention, they were engaged in luminance change detection of this fixation square.

Setup and procedure

Psychophysics experiment

Visual motion stimuli were presented in an area covering a visual angle of 25° width and 25° height. A chinrest at a distance of 55 cm from the monitor was used. Sound was provided via standard stereo headphones. The psychophysical experiment included two sessions of the MOTION, PITCH, and SPEECH experiments i.e., a total of six sessions that were performed on 2 d. Each of the six sessions (lasting 13 min each) included 30 trials of each of the 15 conditions (Fig. 1).

fMRI experiment

As in the psychophysics experiment, visual stimuli were presented in an area covering a visual angle of 25° width and 25° height using a projection screen mounted inside the scanner bore and viewed via a mirror. Sound was provided using MR-compatible headphones (MR Confon) without earplugs at sound pressure levels that enabled effortless discrimination of the sounds. The fMRI experiment included two sessions (lasting 10 min each) of the MOTION, PITCH, and SPEECH experiment, i.e., a total of six sessions that were performed on 2 d. Each session included 44 trials of

each of the six conditions (Fig. 1). Blocks of 12 activation trials were interleaved with 8 s fixation.

Subjects discriminated the direction of the visual motion stimulus and indicated their response as accurately and quickly as possible by a two-choice key press using their dominant hand. The order of the activation conditions was randomized. The order of the sessions, i.e., experimental contexts, was counterbalanced within and across subjects.

Cogent Toolbox (John Romaya, Vision Lab, UCL, London, UK; www.vislab.ucl.ac.uk) for Matlab (MathWorks) was used for stimulus presentation and response recording.

Analysis of behavioral data from psychophysical and functional imaging studies

The behavioral data of the psychophysics and the functional imaging studies were analyzed in an equivalent manner. Separately for the MOTION, PITCH, and SPEECH experiments, percentage of perceived (or judged) upward/rightward motion was calculated for each level of auditory direction signal and visual reliability. A psychometric curve (cumulative Gaussian) was fitted individually for each subject separately for each level of auditory direction signal (i.e., left, absent, or right) (Wichmann and Hill, 2001a,b) (for the psychometric curves of one representative subject, see Fig. 2A). For each subject, the point of subjective equality (PSE) was computed as a measure of the directional bias induced by the auditory direction signal (see Fig. 2B). The PSE is the amount (percentage) of visual direction information for which right/up and left/down judgments are equally likely. To allow generalization to the population level, the subject-specific PSEs were entered into random effects one-way ANOVAs with auditory signal (left, absent, or right) separately for the MOTION, PITCH, and SPEECH experiments. To assess the modulatory effect of context, an additional two-way ANOVA was performed with context (MOTION, PITCH, or SPEECH) and auditory signal (left, absent, or right).

fMRI

A 3 T Siemens Tim Trio system was used to acquire both T1 anatomical volume images and T2*-weighted axial echoplanar images with blood oxygenation level-dependent (BOLD) contrast (GE-EPI; Cartesian k-space sampling; echo time, 40 ms; repetition time, 3080 ms; 38 axial slices, acquired sequentially in ascending direction; matrix, 64×64 ; spatial resolution $3 \times 3 \times 3$ mm³ voxels; interslice gap, 0.4 mm; slice thickness, 2.6 mm). There were two sessions of each experiment (MOTION, PITCH, and SPEECH) with a total of 199 volume images per session. In addition, a motion localizer session with 117 volume images was acquired. The first two volumes of each session were discarded to allow for T1 equilibration effects.

The data were analyzed with statistic parametric mapping [using SPM2 software from the Wellcome Department of Imaging Neuroscience, London, UK; <http://www.fil.ion.ucl.ac.uk/spm> (Friston et al., 1995)]. Scans from each subject were realigned using the first as a reference, spatially normalized into Montreal Neurological Institute standard space (Evans et al., 1992), resampled to $3 \times 3 \times 3$ mm³ voxels and spatially smoothed with a Gaussian kernel of 8 mm full-width half-maximum. The time series in each voxel was high-pass filtered to $1/128$ Hz and globally normalized with proportional scaling. The fMRI experiment was modeled in an event-related manner using regressors obtained by convolving each event-related unit impulse with a canonical hemodynamic response function and its first temporal derivative. The statistical model included the six conditions in our 2×3 factorial design for each of the three contexts (MOTION, PITCH, and SPEECH). Nuisance covariates included the realignment parameters (to account for residual motion artifacts). For further characterization of the data, a second analysis was performed in which the trials of the three visually unreliable conditions in each context were separated into correct and incorrect trials resulting in a 3×3 design. Because the visually reliable conditions elicited only very few or no error trials, these conditions were not modeled as separate regressors for correct and incorrect trials. For both analyses, condition-specific effects for each subject were estimated according to the general linear model. Contrast images (summed over the two sessions

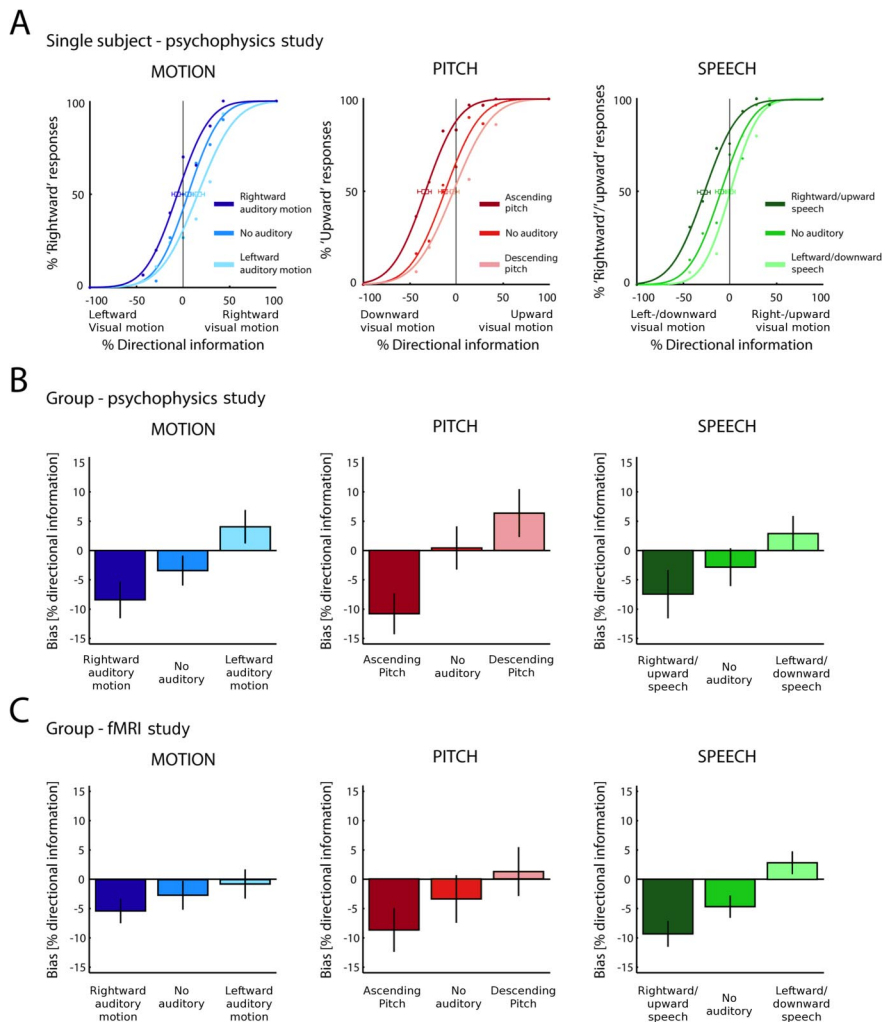


Figure 2. Behavioral biases induced by the different auditory direction signals. **A**, Psychometric functions with PSE of a representative subject from the psychophysics study were obtained for upward/rightward, no auditory, and downward/leftward conditions in the MOTION, PITCH, and SPEECH experiments. Error bars indicate 95% confidence intervals. **B**, **C**, Across-subjects mean of the PSE as a measure for directional bias in the psychophysics (**B**) and fMRI study (**C**). Error bars indicate \pm SEM. The directional bias was significant and statistically indistinguishable across MOTION, PITCH, and SPEECH experiments.

for each of the three contexts) were created for each subject and entered into a second-level one-sample *t* test. The interaction $[(AV_{\text{unreliable}} - V_{\text{unreliable}}) > (AV_{\text{reliable}} - V_{\text{reliable}})]$ and the directed auditory influence (incongruent > congruent and congruent > incongruent) were tested separately for the MOTION, PITCH, and SPEECH experiments. In addition, for comparing effects across contexts, the first-level contrasts were entered into a second-level ANOVA.

Inferences were made at the second level to allow a random effects analysis and inferences at the population level (Friston et al., 1999).

Search volume constraints

To focus on the neural systems that are engaged in audiovisual processing, we limited the search space to the cerebral voxels activated for all activation conditions (within the respective context: MOTION, PITCH, or SPEECH) relative to fixation at a threshold of $p < 0.001$ uncorrected, spatial extent > 100 voxels. In addition, a region of interest-based analysis was performed for left and for right hMT+/V5+ using a sphere of 10 mm radius centered on the group activation peaks [left, (−42, −72, 3) and right, (42, −72, 6)] estimated in the independent motion localizer. These group activation peaks of left and right hMT+/V5+ were objectively determined as the main peaks within bounding boxes centered on hMT+/V5+ coordinates with a $\frac{1}{2}$ diameter equal to 2 SDs as reported by Dumoulin et al. (2000) (right, $x = 44 \pm 3.3$, $y = -67 \pm 3.1$, $z = 0 \pm 5.1$; left, $x = -47 \pm 3.8$, $y = -76 \pm 4.9$, $z = 2 \pm 2.7$).

Unless otherwise stated, we report activations at $p < 0.05$ corrected at the cluster level for multiple comparisons in the volume of interest using an auxiliary (uncorrected) voxel threshold of $p < 0.01$. This auxiliary threshold defines the spatial extent of activated clusters, which form the basis of our (corrected) inference.

Results of the random effects analysis are superimposed onto a T1-weighted brain that was generated by averaging the normalized brains of the 21 subjects, using MRIcro software (<http://www.sph.sc.edu/comd/rorden/micro.html>).

Results

In a series of three experiments, we investigated the influence of three types of auditory direction signals on visual motion discrimination: natural auditory MOTION, metaphoric PITCH, and linguistic SPEECH signals. Each experiment conformed to a factorial design manipulating (1) reliability of the visual motion direction (five or two levels) and (2) the auditory direction signal (three levels: directionally congruent, absent, or incongruent to the visual signal). In a visual selective attention paradigm, subjects discriminated the direction of high-resolution apparent visual motion (Fig. 1).

We first characterized subjects' performance thoroughly in a psychophysics experiment. Careful stimulus adjustment enabled us to equate MOTION, PITCH, and SPEECH direction signals in terms of their bias on perceived direction of visual motion.

In a more constrained fMRI study, we then investigated whether auditory MOTION, PITCH, and SPEECH stimuli modulate and interact with visual motion processing at different levels of the cortical hierarchy.

Psychophysical experiment

Figure 2A shows the psychometric curves for one representative subject, obtained in the three different auditory conditions. For directional bias as measured by the shift of the point of subjective equality (Fig. 2A, B), one-way ANOVAs computed separately for each context with the factor auditory direction signal (left, absent, and right), revealed a main effect of auditory direction signal in the MOTION ($F_{(1,2,13.8)} = 14.0$; $p < 0.002$), PITCH ($F_{(1,2,14.1)} = 34.5$; $p < 0.001$), and SPEECH ($F_{(1,6,18.7)} = 10.5$; $p < 0.002$) experiments. To evaluate the effect of context, a two-way ANOVA with context (MOTION, PITCH, and SPEECH) and auditory direction signal (three levels) was performed. This ANOVA showed a main effect of auditory direction signal ($F_{(1,3,15.1)} = 39.6$; $p < 0.001$) but no main effect of context and crucially no interaction between context and auditory direction signal, indicating a comparable bias across the three contexts. Similarly, analysis of accuracy and reaction times in a three-way ANOVA with context (MOTION, PITCH, and SPEECH), visual reliability (four levels), and auditory direction signal (congruent vs absent/unimodal vs incongruent) replicated the main effect of auditory direction signal. The interaction between visual reliabil-

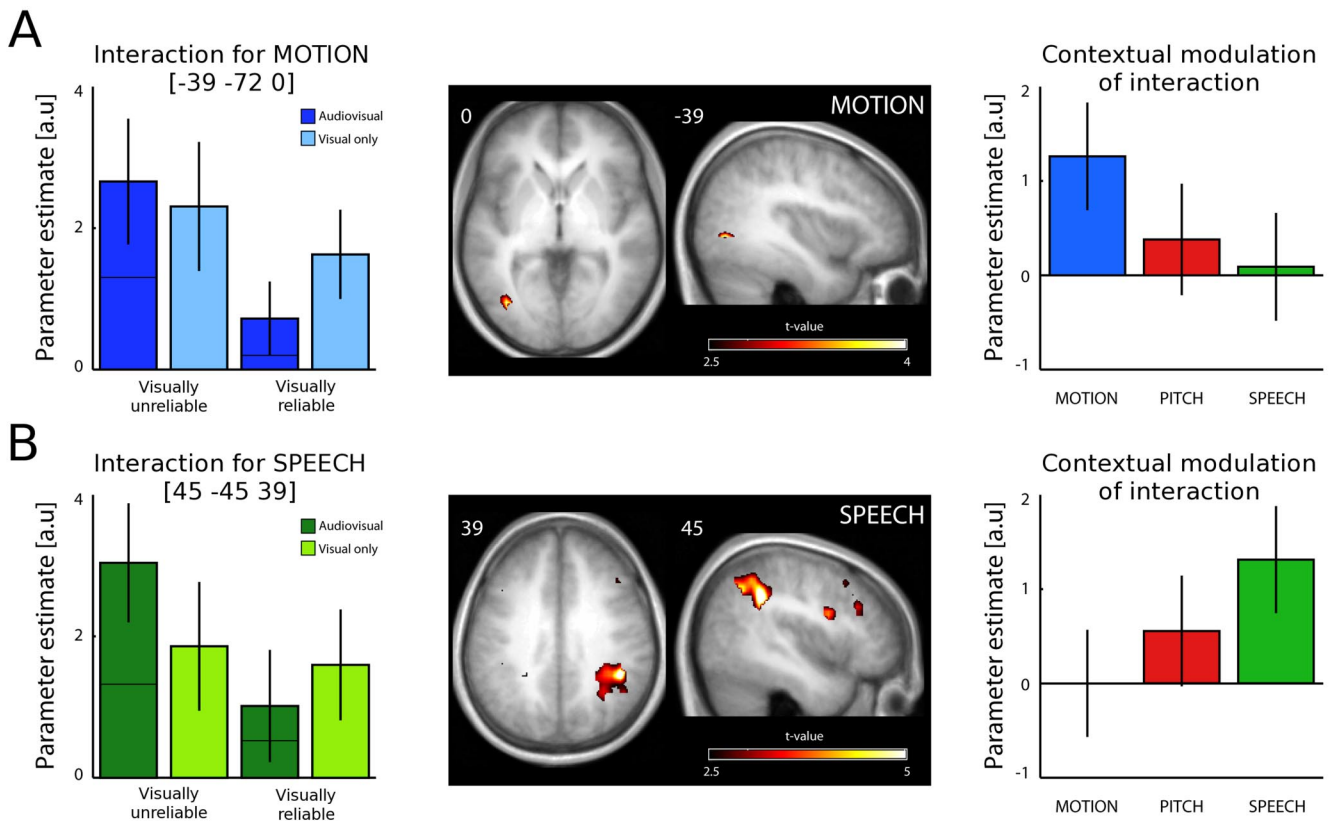


Figure 3. Influence of auditory direction signals on visual motion processing in the MOTION (**A**) and SPEECH (**B**) experiments. Left, Parameter estimates for $AV_{\text{unreliable}}$, $V_{\text{unreliable}}$, AV_{reliable} , and V_{reliable} for the MOTION experiment in the left hMT+/V5+ [(-39, -72, 0)] (**A**) and the SPEECH experiment in the right IPS [(45, -45, 39)] (**B**). The bars for the bimodal conditions represent combined estimates from congruent (bottom part of the bars) and incongruent (top part of the bars) conditions. Error bars indicate 90% confidence intervals. Middle, The activations pertaining to the audiovisual interaction ($AV_{\text{unreliable}} - V_{\text{unreliable}} > AV_{\text{reliable}} - V_{\text{reliable}}$) for MOTION (**A**) and SPEECH (**B**) are displayed on axial and sagittal slices of a mean EPI image created by averaging the subjects' normalized echo planar images (height threshold, $p < 0.01$; spatial extent > 0 voxels; see Materials and Methods). Right, Contextual modulation of the AV interaction: parameter estimates of the audiovisual interaction effect in MOTION (blue), PITCH (red), and SPEECH (green) are displayed for the left hMT+/V5+ [(-39, -72, 0)] (**A**) and the right IPS [(45, -45, 39)] (**B**). In the left hMT+/V5+, the interaction effect gradually decreases for MOTION $>$ PITCH $>$ SPEECH; in the right IPS, it gradually increases for MOTION $<$ PITCH $<$ SPEECH. Error bars indicate 90% confidence intervals.

ity and auditory direction signal and the absence of a contextual modulation of the congruency effect indicated comparable magnitude of the auditory influence on the perceived direction of visual motion across contexts (for full characterization of behavioral data, see supplemental Results and table, available at www.jneurosci.org as supplemental material).

Behavior during fMRI experiment

For directional bias (Fig. 2C), one-way ANOVAs with auditory direction signal (left, absent, and right) revealed a main effect of audition in the MOTION ($F_{(2,40)} = 4.4$; $p < 0.018$), PITCH ($F_{(1.5,30.1)} = 5.2$; $p < 0.018$), and SPEECH ($F_{(1.6,31.26)} = 19.1$; $p < 0.001$) experiments. A two-way ANOVA with context (MOTION, PITCH, and SPEECH) and auditory direction signal (three levels) showed a main effect of auditory direction signal ($F_{(1.6,32.3)} = 17.9$; $p < 0.001$) but no main effect of context and importantly no interaction between context and auditory direction signal. In line with the psychophysics results, this demonstrates that MOTION, PITCH, and SPEECH auditory direction signals induce a comparable bias on perceived direction of visual motion. Additional analyses of performance accuracy and reaction times of the fMRI behavioral data led to equivalent conclusions (for full characterization of behavioral data, see supplemental Results and table, available at www.jneurosci.org as supplemental material).

To summarize, both psychophysics and behavioral data acquired during fMRI scanning demonstrate that auditory direction signals induce a significant bias on perceived direction of visual motion that is comparable for the MOTION, PITCH, and SPEECH contexts. This is an important finding for two reasons. First, it demonstrates that through careful stimulus adjustment, natural, metaphoric, and linguistic auditory direction signals similarly influence visual motion discrimination, although vision is usually considered the dominant modality in motion perception. Second, it allows us to investigate whether different neural systems mediate the influence of different classes of auditory direction signals unconfounded by behavioral differences.

Functional imaging results

Influence of auditory direction signals on visual motion processing
For each context, we identified regions in which auditory direction signals influence visual motion processing regardless of the directional congruency of the two signals. Behaviorally, the task-irrelevant auditory stimulus exerted a stronger influence and even disambiguated the direction of the visual stimulus, when the visual motion information was unreliable. An audiovisual enhancement for the bimodal (AV) relative to the unimodal visual (V) response was therefore expected primarily when the visual stimulus was unreliable. Hence, we tested for the interaction

Table 1. Influence of auditory direction signals on visual motion processing

Region		Coordinates	z-score at peak	Number of voxels	<i>p</i> corrected for spatial extent
$(AV_{\text{unreliable}} - V_{\text{unreliable}}) > (AV_{\text{reliable}} - V_{\text{reliable}})$ (separately for the MOTION, PITCH, and SPEECH contexts)					
Motion					
hMT+/V5+	Left	−39, −72, 0	3.34	29	0.030 VOI ₂
Planum temporale	Right (correct trials only)	57, −24, 15	3.18	75	0.042 VOI ₁
	Left	−45, −36, 12	3.59	49	*
Posterior middle temporal gyrus	Right (correct trials only)	42, −57, 12	3.63	71	0.050 VOI ₁
	Left	−42, −60, 6	3.27	27	*
Pitch					
Superior parietal cortex	Left	−36, −48, 69	3.22	29	*
	Right	57, −51, 51	2.84	25	*
Speech					
Intraparietal sulcus	Right	45, −45, 39	4.74	302	0.001 VOI ₁
	Left	−42, −54, 45	2.67	13	*
Ventral premotor cortex	Right	42, 6, 27	3.68	37	*
Inferior frontal sulcus	Right	42, 27, 30	3.00	29	*
Effect of congruency between auditory and visual direction signals (incongruent > congruent)					
Speech					
Posterior superior temporal gyrus	Left	−66, −39, 18	4.67	88	0.036 VOI ₁
Inferior frontal sulcus	Right	48, 18, 39	4.48	95	0.027 VOI ₁

VOI₁, Activations for (all conditions > fixation) separately for MOTION, PITCH, or SPEECH experiments; *p* < 0.001 uncorrected; spatial extent > 100 voxels. VOI₂, Spherical ROI with radius of 10 mm centered on peak from independent functional MT localizer; * indicates uncorrected

$(AV_{\text{unreliable}} - V_{\text{unreliable}}) > (AV_{\text{reliable}} - V_{\text{reliable}})$ separately for the MOTION, PITCH, and SPEECH experiments.

For the MOTION experiment (Fig. 3A, Table 1), we observed a significant interaction effect in left hMT+/V5+ within the spherical mask defined by our independent hMT+/V5+ localizer (see Materials and Methods). When limiting the contrast primarily to correct trials (for details of this additional analysis, see Materials and Methods), significant interactions were observed in three clusters: again, in the left hMT+/V5+ extending into more anterior and superior regions, in the right posterior middle temporal gyrus, and the posterior superior temporal gyrus/planum temporale. Thus, the modulatory effect of the auditory signal seems to be slightly more pronounced when subject's directional judgment is correct, i.e., consistent with the direction of the visual motion (although a direct comparison between the interactions for correct and incorrect trials was not significant). Our results demonstrate that auditory MOTION signals modulate visual motion processing in a set of areas previously implicated in unimodal auditory (Warren et al., 2002) or visual (Zeki et al., 1991) motion processing. More specifically, auditory input amplifies the visual signal when the visual stimulus is unreliable but suppresses visually induced responses when the visual stimulus provides a reliable direction signal. Interestingly, as demonstrated in the parameter estimate plots (Fig. 3A, left), the modulatory effect of the auditory MOTION signals seems to emerge regardless of whether the directions of auditory and visual motion are congruent or incongruent. The absence of differences in hMT+/V5+ activation between congruent and incongruent conditions may be attributable to the use of a visual selective attention task, in which AV (in)congruency can induce several counteracting effects. First, congruent auditory motion may induce an amplification of hMT+/V5+ responses reflecting the emergence of a coherent motion percept when the visual stimulus is unreliable ($AV_{\text{unreliable}}$ congruent). When the visual stimulus is reliable and elicits a strong motion percept by itself (AV_{reliable} congruent), congruent auditory motion may mediate more efficient processing, leading to a suppression of the visual response (see related discussions about neural mechanisms of

priming in the studies of Henson et al., 2000, 2003). Second, spatially or semantically incongruent audiovisual stimuli (AV_{reliable} incongruent) are generally thought to induce a suppression relative to the unimodal response [see neurophysiological results (Meredith and Stein, 1996) or functional imaging of AV speech processing in humans (van Atteveldt et al., 2004)]. Third, in a visual-selective attention paradigm, incongruent, yet irrelevant, auditory motion may induce amplification of task-relevant motion information in hMT+/V5+ via top-down modulation to overcome the interfering auditory motion stimulus (Noppeney et al., 2008). This amplification may be particularly pronounced when the visual stimulus that needs to be discriminated is unreliable ($AV_{\text{unreliable}}$ incongruent). In short, the complex balance between these three effects may determine whether congruent or incongruent conditions induce AV enhancement or suppression relative to the unimodal visual conditions.

In addition to “putatively” unimodal motion processing areas, we expected higher-level association areas such as anterior intraparietal sulcus (IPS) (Bremmer et al., 2001) to be involved in AV motion integration. Surprisingly, the IPS exhibited an audiovisual interaction effect only at a very low threshold of significance that did not meet our stringent statistical criteria [left, (−27, −51, 45), *z*-score of 2.2; and right (36, −42, 45), *z*-score of 2.58].

To further evaluate whether the modulatory effect of audition is selective for natural auditory MOTION signals, we entered the interaction (estimated for MOTION, PITCH, and SPEECH) into an ANOVA. We tested for a linear decrease of the interaction effect across the three contexts: MOTION > PITCH > SPEECH at the peak voxel in left hMT+/V5+ [(−39, −72, 0); *z*-score of 2.27; *p*_{uncorr} = 0.012]. As shown in the parameter estimate plots, the interaction effect was most pronounced for natural auditory MOTION. Hence, the modulatory effect of audition on the mean BOLD response of hMT+/V5+ is selective for natural auditory MOTION signals.

For the PITCH experiment, no significant interactions were observed (even when limiting the analysis to correct trials only). For complete characterization of the data, we report the bilateral

superior parietal/intraparietal clusters that were observed at an uncorrected level of significance. These activation foci were centered more dorsally but partly overlapping with those found for SPEECH (Table 1, Fig. 4).

For the SPEECH experiment (Fig. 3*B*, Table 1), a highly significant interaction was detected in the right anterior IPS [(45, -45, 39)] and at an uncorrected level in the right ventral premotor cortex [(42, 6, 27)], which have been previously implicated in polymodal motion processing [e.g., compare activation peaks reported by Bremmer et al. (2001): right putative ventral intraparietal area (VIP), (38, -44, 46); right ventral premotor, (52, 10, 30)]. As in the case of auditory motion, the modulatory effect of auditory SPEECH signals emerged again regardless of whether the directions of auditory and visual motion were congruent or incongruent (see parameter estimate plots in Fig. 3*B*).

Furthermore, the modulatory effect of audition is selective for SPEECH signals as indicated by a significant linear decrease of the interaction effects across the three contexts: SPEECH > PITCH > MOTION at the IPS peak voxel [(45, -45, 39); peak *z*-score of 3.17; $p_{\text{uncorr}} = 0.001$].

As reported above, a gradual contextual modulation of the interaction effect was found at the peak coordinates in hMT+/V5+ and right IPS identified through the AV interaction for MOTION and SPEECH, respectively. To further characterize the spatial specificity of this finding in an unbiased manner, we defined a search mask by identifying all regions that showed increased activation for the interaction including all three contexts (MOTION, PITCH, and SPEECH) using a liberal threshold ($p < 0.05$ uncorrected, spatial extent >50 voxels). A general interaction effect was observed in five regions: left hMT+/V5+, left and right IPS, right superior temporal gyrus, and right ventral premotor cortex. We then tested at the peak coordinates of each of these five regions for a linear decrease of the interaction effect according to (1) MOTION > PITCH > SPEECH and (2) SPEECH > PITCH > MOTION (note that the general interaction effect is orthogonal to the contextual modulation of the interaction, i.e., three-way interaction, thus enabling an unbiased evaluation of our contrast of interest). After Bonferroni's correction for multiple comparisons (i.e., number of regions), the only regions showing a significant gradual decrease or increase in activation were left hMT+/V5+ [(-39, -72, 0); *z*-score of 2.27; $p < 0.012$] for MOTION > PITCH > SPEECH and right IPS [(48, -45, 45); *z*-score of 2.54; $p < 0.006$] for SPEECH > PITCH > MOTION (Fig. 3*A, B*, right side).

For completeness, we also tested for the reverse interaction ($AV_{\text{reliable}} - V_{\text{reliable}} > AV_{\text{unreliable}} - V_{\text{unreliable}}$) separately for each context but did not observe any significant effects.

In summary, we found a double dissociation of auditory influence on visual motion processing across the three contexts: whereas natural auditory MOTION direction signals influenced visual motion processing primarily in hMT+/V5+ and related areas previously implicated in auditory motion processing (i.e., planum temporale and anterior middle temporal), the effect of SPEECH direction signals was observed in a higher-level frontoparietal system. Interestingly, in both cases, the magnitude of the PITCH interaction effect was between that of SPEECH and MOTION. Although the influence of PITCH direction signals was behaviorally indistinguishable from MOTION and SPEECH, we

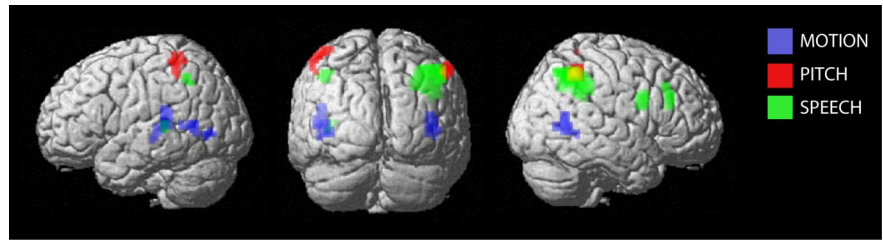


Figure 4. Overview of the neural systems mediating the influence of naturalistic, metaphoric, and linguistic auditory direction signals on visual motion processing. The audiovisual interaction effects for MOTION (blue), PITCH (red), and SPEECH (green) are rendered on a template of the whole brain (height threshold, $p < 0.01$ uncorrected; spatial extent >10 voxels; see Materials and Methods).

observed a modulatory effect selective for PITCH direction signals only at a lower level of significance in bilateral anterior superior parietal lobes (for overview, see summary Fig. 4).

Effect of congruency between auditory and visual direction signals

Next, we evaluated separately for each context whether the congruency between visual and auditory direction signals influences audiovisual motion processing. Because the direction signal in the visual modality (and hence a meaningful AV congruency relationship) was nearly completely removed in the visually unreliable conditions (as indicated by only 60% correct responses), we limited this contrast to the visually reliable trials only. Although in classical multisensory integration paradigms (e.g., passive contexts or bimodal attention) increased activations are found for congruent relative to incongruent trials, quite the opposite pattern (i.e., increased activation for incongruent relative to congruent trials) is observed in selective attention paradigms (Weissman et al., 2004; Noppeney et al., 2008). These incongruency effects may represent a prediction error signal (Rao and Ballard, 1999; Garrido et al., 2007) or an attempt to amplify the task-relevant visual signal to overcome the interfering auditory signal. Hence, in our experiment, we expected to find increased activation in task-relevant visual motion areas as well as higher control areas.

Contrary to our expectations, we did not observe increased hMT+/V5+ responses for incongruent relative to congruent trials in any of the three contexts, nor did we observe any incongruency effects for the MOTION and PITCH contexts elsewhere in the brain. However, for SPEECH context, we observed increased activation for incongruent relative to congruent trials in the right inferior frontal sulcus and in the left posterior superior temporal gyrus (Fig. 5, Table 1), which was shown to be selective for the speech context when the incongruency effects were compared directly across contexts. These results are in line with numerous studies of cognitive control implicating the dorsolateral prefrontal cortex in dynamic selection of task-relevant information, by biasing processing in sensory areas, particularly in conflicting contexts (MacDonald et al., 2000; Miller 2000; Matsumoto and Tanaka, 2004; Noppeney et al., 2008).

For completeness, we also tested the opposite statistical comparison (i.e., congruent > incongruent), which revealed activation in the right supramarginal gyrus for the PITCH experiment only [(48, -39, 30); *z*-score of 3.94; voxels, 161; $p = 0.001$].

Lateralization of the AV interactions or incongruency effects

To investigate whether any of the reported effects were significantly lateralized, a hemisphere by condition interaction analysis was performed (for additional details, see supplemental data, available at www.jneurosci.org as supplemental material). This analysis demonstrated that none of the reported interaction or incongruency effects were significantly lateralized (after correc-

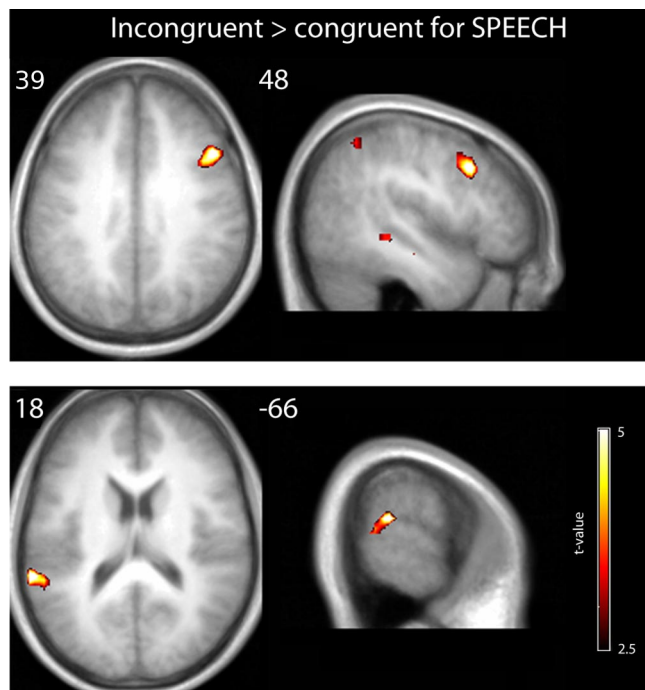


Figure 5. Increased activations for incongruent relative to congruent visuoauditory stimuli in the SPEECH experiment are shown on axial and coronal slices of a mean echo planar image created by averaging the subjects' normalized echo planar images (height threshold, $p < 0.01$; spatial extent > 0 voxels; see Materials and Methods).

tion for multiple comparisons). To characterize the bilateral nature of the effects, Table 1 also reports the corresponding interaction and incongruency effects in the homologous regions of the contralateral hemisphere at an uncorrected threshold of significance and spatial extent > 10 voxels (equivalent to the thresholds applied in Fig. 4).

Discussion

The present study demonstrates that, although auditory MOTION, PITCH, and SPEECH signals behaviorally induce a comparable directional bias, their influences emerge at different levels of the cortical hierarchy. Auditory MOTION signals interact with visual input in visual and auditory motion areas, SPEECH direction signals in a higher-level frontoparietal system.

With the help of externalized virtual-space sounds (Pavani et al., 2002), a high spatial resolution (i.e., 14 flash positions), strict audiovisual colocalization through combined binaural (i.e., interaural time and amplitude differences), and subject-specific monoaural filtering cues, this study shows a reliable directional bias for auditory MOTION on visual motion discrimination. In contrast, previous studies deprived of ecologically valid AV motion cues or using apparent audiovisual motion stimuli with a low spatial resolution have failed to show an influence of auditory motion on visual motion perception (Soto-Faraco et al., 2004) or on activations in hMT+/V5+ (Baumann and Greenlee, 2007) emphasizing the importance of a natural experimental set up for studying AV integration (cf. Meyer et al., 2005). Furthermore, our results replicate and extend the metaphoric auditory capture of dynamic PITCH to spatiotemporally confined visual stimuli (Maeda et al., 2004) and demonstrate that even spoken words (e.g., left/right) with only arbitrary relationships to physical dimensions induced a reliable directional bias. Crucially, a comparable directional bias was obtained for all three classes of direction signals even when stimulus length and audiovisual timing

relations were equivalent. This allowed us to investigate whether natural, metaphoric, and linguistic AV motion interactions are mediated by distinct neural systems unconfounded by differences in (1) temporal parameters and stimulus length and/or (2) behavioral differences as measured in terms of bias, accuracy, and reaction times.

Indeed, despite controlling for stimulus parameters and behavioral performance measures, auditory MOTION and SPEECH signals influenced neural processes underlying visual motion discrimination at distinct levels of the cortical hierarchy. Auditory MOTION influences emerged in auditory–visual motion areas encompassing both (1) “classical” visual motion area hMT+/V5+ and (2) the planum temporale and a middle temporal region that are both thought to be part of the posterior auditory stream of spatial (Rauschecker and Tian, 2000; Warren and Griffiths, 2003) and motion (region anterior to hMT+/V5+) (Warren et al., 2002) processing. Interestingly, the activation pattern sharply contrasts with that observed in dynamic visual capture, in which intraparietal activation is followed by an activation increase in hMT+/V5+ but decrease in auditory motion regions (Alink et al., 2008). Thus, in dynamic visual capture, parietal regions seem to send top-down biasing signals that see-saw between auditory and visual motion activations. In contrast, our interaction effect emerges simultaneously in auditory and visual motion areas in the absence of significant interactions in higher-order intraparietal areas. This suggests that auditory and visual information may become integrated through direct interactions between auditory and visual regions rather than convergence in parietal association areas [see related accounts of early multisensory integration (Schroeder and Foxe, 2002, 2005; Ghazanfar et al., 2005; Kayser et al., 2005; Ghazanfar and Schroeder, 2006)].

Auditory SPEECH signals induced the opposite pattern with an interaction effect in the anterior IPS and, at an uncorrected level of significance, in the premotor cortex but not in AV motion areas. The ventral premotor–parietal system has been implicated previously in polymodal motion processing on the basis of conjunction analyses (Bremmer et al., 2001) and proposed as the putative human homolog of macaque VIP–ventral premotor circuitry (Cooke et al., 2003; Bremmer, 2005; Rushworth et al., 2006; Sereno and Huang, 2006) (but see Bartels et al., 2008). Furthermore, anterior to classical visuospatial attention areas (Corbetta and Shulman, 1998; Kastner et al., 1999; Corbetta and Shulman, 2002), these anterior parietal areas have also been implicated in crossmodal attention and top-down control (Macaluso et al., 2000; Macaluso et al., 2003; Shomstein and Yantis, 2004; Townsend et al., 2006). Thus, spatial language influences motion discrimination at higher-order association regions to guide perceptual decisions in our multisensory dynamic environment.

A direct comparison of the AV interaction effects across contexts confirmed this double dissociation: whereas AV motion areas showed a gradual activation decrease for MOTION $>$ PITCH $>$ SPEECH, the IPS showed an activation increase for MOTION $<$ PITCH $<$ SPEECH. Interestingly, PITCH is associated with intermediate activation levels in both areas. These results may reflect the “ambivalent” nature of metaphoric interactions. On the one hand, infants even within their first year associate visual arrows pointing up or down with rising or falling pitch (Wagner et al., 1981), suggesting intrinsic similarities between pitch and motion dimensions (Melara and O'Brien, 1987; Maeda et al., 2004; Marks, 2004). However, given our cultural education throughout lifespan, metaphoric interactions may well

become incorporated in language and be mediated at least in part by semantic or linguistic processes in adulthood. PITCH-selective interactions were found at an uncorrected threshold in bilateral superior/intraparietal regions that are partly overlapping with the SPEECH effects (Fig. 4). In addition to its well established role in spatial processing (Ungerleider and Mishkin, 1982; Culham and Kanwisher, 2001), the superior parietal cortex has been implicated previously in pitch-sequence analysis (Griffiths et al., 1999) and pitch memory (Zatorre et al., 1994; Gaab et al., 2003). Additional evidence for a link between spatial and pitch processing comes from studies of amusia, showing a correlation between deficits in pitch direction analysis and impaired spatial processing (Douglas and Bilkey, 2007). In light of these studies, our results may point to the superior parietal lobe as a site for interactions between pitch and visual motion within a shared spatial reference frame.

Additional characterization of the AV interaction (see parameter estimate plots in Fig. 3) revealed that the auditory stimulus amplified the response to visually unreliable but suppressed the response to visually reliable stimuli. Because the unimodal visually unreliable stimulus elicits a greater response than the reliable stimulus (most likely attributable to attentional top-down modulation), this means that paradoxically auditory input amplifies the stronger BOLD response and attenuates the weaker response to visual input. This response pattern demonstrates the context dependency of the classical law of inverse effectiveness whereby multisensory enhancement is supposedly greatest for the least effective (smaller response) stimulus. Instead, our results suggest that auditory input may disambiguate degraded and unreliable stimuli through multisensory response enhancement, possibly using similar mechanisms as priming-induced activation increases that mediate the recognition of degraded objects that were not recognized during initial presentation (Dolan et al., 1997; Henson et al., 2000; Henson, 2003).

In conclusion, although natural, metaphoric, and linguistic auditory direction signals induce a comparable directional bias on motion discrimination, they emerge at different levels of the cortical hierarchy. While natural auditory MOTION modulates activations in auditory and visual motion areas, SPEECH direction signals influence an anterior intraparietal–ventral premotor circuitry that has been implicated previously in polymodal motion processing and multisensory control. From a cognitive perspective, this double dissociation suggests that natural and linguistic biases emerge at perceptual and decisional levels, respectively. From the perspective of neural processing, to our knowledge, this is the first compelling demonstration in humans that natural auditory and visual motion information may be integrated through direct interactions between auditory and visual motion areas. In contrast, SPEECH direction signals that are culturally learnt influence primarily higher-level convergence and control regions.

References

Alais D, Burr D (2004) No direction-specific bimodal facilitation for audio-visual motion detection. *Brain Res Cogn Brain Res* 19:185–194.

Alink A, Singer W, Muckli L (2008) Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J Neurosci* 28:2690–2697.

Bartels A, Zeki S, Logothetis NK (2008) Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb Cortex* 18:705–717.

Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. *Cereb Cortex* 17:1433–1443.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and

anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex* 17:2172–2189.

Bremmer F (2005) Navigation in space: the role of the macaque ventral intraparietal area. *J Physiol* 566:29–35.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fmri study strongly implies equivalencies between humans and monkeys. *Neuron* 29:287–296.

Cooke DF, Taylor CS, Moore T, Graziano MS (2003) Complex movements evoked by microstimulation of the ventral intraparietal area. *Proc Natl Acad Sci U S A* 100:6163–6168.

Corbetta M, Shulman GL (1998) Human cortical mechanisms of visual attention during orienting and search. *Philos Trans R Soc Lond B Biol Sci* 353:1353–1362.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.

Culham JC, Kanwisher NG (2001) Neuroimaging of cognitive functions in human parietal cortex. *Curr Opin Neurobiol* 11:157–163.

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RS, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. *Nature* 389:596–599.

Douglas KM, Bilkey DK (2007) Amusia is associated with deficits in spatial processing. *Nat Neurosci* 10:915–921.

Dumoulin SO, Bittar RG, Kabani NJ, Baker CL Jr, Le Goualher G, Bruce Pike G, Evans AC (2000) A new anatomical landmark for reliable identification of human area v5/mt: a quantitative analysis of sulcal patterning. *Cereb Cortex* 10:454–463.

Evans AC, Collins DL, Milner B (1992) An mri-based stereotactic atlas from 250 young normal subjects. *Soc Neurosci Abstr* 18:408.

Friston K, Holmes A, Worsley K, Poline JB, Frith C, Frackowiak R (1995) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210.

Friston KJ, Holmes AP, Price CJ, Büchel C, Worsley KJ (1999) Multisubject fmri studies and conjunction analyses. *Neuroimage* 10:385–396.

Gaab N, Gaser C, Zaehle T, Jancke L, Schlaug G (2003) Functional anatomy of pitch memory: an fmri study with sparse temporal sampling. *Neuroimage* 19:1417–1426.

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. *Proc Natl Acad Sci U S A* 104:20961–20966.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285.

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012.

Griffiths TD, Johnsrude I, Dean JL, Green GG (1999) A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport* 10:3825–3830.

Henson RN (2003) Neuroimaging studies of priming. *Prog Neurobiol* 70:53–81.

Henson R, Shallice T, Dolan R (2000) Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287:1269–1272.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron* 22:751–761.

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48:373–384.

Krekelberg B, Vatakis A, Kourtzi Z (2005) Implied motion from form in the human visual cortex. *J Neurophysiol* 94:4373–4386.

Macaluso E, Frith CD, Driver J (2000) Modulation of human visual cortex by crossmodal spatial attention. *Science* 289:1206–1208.

Macaluso E, Eimer M, Frith CD, Driver J (2003) Preparatory states in cross-modal spatial attention: spatial specificity and possible control mechanisms. *Exp Brain Res* 149:62–74.

MacDonald AW 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288:1835–1838.

Maeda F, Kanai R, Shimojo S (2004) Changing pitch induced visual motion illusion. *Curr Biol* 14:R990–R991.

Marks L (2004) Cross-modal interactions in speeded classification. In: *The handbook of multisensory processes*, pp 85–105. Cambridge, MA: MIT.

Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli

- RA (2007) Multisensory interactions within human primary cortices revealed by bold dynamics. *Cereb Cortex* 17:1672–1679.
- Mateeff S, Hohnsbein J, Noack T (1985) Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception* 14:721–727.
- Matsumoto K, Tanaka K (2004) Neuroscience. Conflict and cognitive control. *Science* 303:969–970.
- Melara R, O'Brien T (1987) Interaction between synesthetically corresponding dimensions. *J Exp Psychol* 116:323–336.
- Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol* 75:1843–1857.
- Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals. *Neuroreport* 12:2557–2560.
- Meyer GF, Wuerger SM, Röhrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Exp Brain Res* 166:538–547.
- Miller EK (2000) The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 1:59–65.
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609.
- Pavani F, Macaluso E, Warren JD, Driver J, Griffiths TD (2002) A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr Biol* 12:1584–1590.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806.
- Rushworth MF, Behrens TE, Johansen-Berg H (2006) Connection patterns distinguish 3 regions of human parietal cortex. *Cereb Cortex* 16:1418–1430.
- Saberi K, Perrott DR (1990) Minimum audible movement angles as a function of sound source trajectory. *J Acoust Soc Am* 88:2639–2644.
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol* 15:454–458.
- Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res* 14:187–198.
- Sereno MI, Huang RS (2006) A human parietal face area contains aligned head-centered visual and tactile maps. *Nat Neurosci* 9:1337–1343.
- Shomstein S, Yantis S (2004) Control of attention shifts between vision and audition in human cortex. *J Neurosci* 24:10702–10706.
- 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 (2004) Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *J Exp Psychol Hum Percept Perform* 30:330–345.
- Townsend J, Adamo M, Haist F (2006) Changing channels: an fmri study of aging and cross-modal attention shifts. *Neuroimage* 31:1682–1692.
- Ungerleider LG, Mishkin M (1982) Two visual systems. In: *Analysis of visual behavior* (Ingle DJ, Goodale MA, Mansfield RJW, eds), pp 549–586. Cambridge, MA: MIT.
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- Wagner S, Winner E, Cicchetti D, Gardner H (1981) “Metaphorical” mapping in human infants. *Child Dev* 52:728–731.
- Warren JD, Griffiths TD (2003) Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J Neurosci* 23:5799–5804.
- Warren JD, Zielinski BA, Green GG, Rauschecker JP, Griffiths TD (2002) Perception of sound-source motion by the human brain. *Neuron* 34:139–148.
- Weissman DH, Warner LM, Woldorff MG (2004) The neural mechanisms for minimizing cross-modal distraction. *J Neurosci* 24:10941–10949.
- Wichmann FA, Hill NJ (2001a) The psychometric function. I. Fitting, sampling, and goodness of fit. *Percept Psychophys* 63:1293–1313.
- Wichmann FA, Hill NJ (2001b) The psychometric function. II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* 63:1314–1329.
- Zatorre RJ, Evans AC, Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. *J Neurosci* 11:641–649.