

Sensory and Striatal Areas Integrate Auditory and Visual Signals into Behavioral Benefits during Motion Discrimination

Sebastian von Saldern¹ and Uta Noppeney^{1,2}

¹Max-Planck-Institute for Biological Cybernetics, 72076 Tübingen, Germany, and ²Centre for Computational Neuroscience and Cognitive Robotics, University of Birmingham, Birmingham B15 2TT, United Kingdom

For effective interactions with our dynamic environment, it is critical for the brain to integrate motion information from the visual and auditory senses. Combining fMRI and psychophysics, this study investigated how the human brain integrates auditory and visual motion into benefits in motion discrimination. Subjects discriminated the motion direction of audiovisual stimuli that contained directional motion signal in the auditory, visual, audiovisual, or no modality at two levels of signal reliability. Therefore, this $2 \times 2 \times 2$ factorial design manipulated: (1) auditory motion information (signal vs noise), (2) visual motion information (signal vs noise), and (3) reliability of motion signal (intact vs degraded). Behaviorally, subjects benefited significantly from audiovisual integration primarily for degraded auditory and visual motion signals while obtaining near ceiling performance for “unisensory” signals when these were reliable and intact. At the neural level, we show audiovisual motion integration bilaterally in the visual motion areas hMT+/V5+ and implicate the posterior superior temporal gyrus/planum temporale in auditory motion processing. Moreover, we show that the putamen integrates audiovisual signals into more accurate motion discrimination responses. Our results suggest audiovisual integration processes at both the sensory and response selection levels. In all of these regions, the operational profile of audiovisual integration followed the principle of inverse effectiveness, in which audiovisual response suppression for intact stimuli turns into response enhancements for degraded stimuli. This response profile parallels behavioral indices of audiovisual integration, in which subjects benefit significantly from audiovisual integration only for the degraded conditions.

Introduction

Integration of motion information from the auditory and visual senses is critical for an organism’s survival, enabling faster and more accurate responses to our dynamic environment. The adaptive advantage of integrating motion information across the senses is most prominent when individual motion cues are unreliable. For example, in a dense forest, we may more easily detect a camouflaged animal and discriminate the direction in which it is heading by combining the degraded visual input with the noise of its footsteps.

From a behavioral perspective, it is controversial whether audiovisual benefits emerge at sensory, perceptual, or decisional levels. In support of perceptual or decisional integration, psychophysics studies demonstrated that the increase in motion detection for audiovisual relative to “unisensory” stimulation can be explained by maximum likelihood estimation or probability

summation (Wuerger et al., 2003; Alais and Burr, 2004). Conversely, in support of sensory integration, directionally congruent motion sounds were shown to help observers decide which of two intervals contained motion, even when the sounds were presented in both intervals and thus were uninformative for discrimination performance (Kim et al., 2012).

At the neural level, multisensory integration is thought to emerge in a widespread neural system including subcortical, primary sensory, and higher-order association areas (Ghazanfar and Schroeder, 2006). Focusing on motion processing, functional imaging in humans demonstrated that visual, auditory, and tactile motion converges in a parietal area, considered a putative homolog of macaque area VIP (Bremmer et al., 2001). Initial evidence also suggests audiovisual interplay in visual and auditory areas (Baumann and Greenlee, 2007). Under auditory selective attention, visual motion capture is associated with activation increases in hMT+/V5+ and decreases in auditory cortex (Alink et al., 2008). Conversely, under visual selective attention, motion sounds were shown to amplify the hMT+/V5+ response to concurrent visual motion (Sadaghiani et al., 2009).

These results indicate that audiovisual influences on motion processing are bidirectional, operating from vision to audition and vice versa. However, because previous studies did not include audiovisual, auditory, and visual conditions in one experiment, they cannot address the critical question of how audiovisual integration facilitates motion discrimination relative to unisensory

Received June 24, 2012; revised April 2, 2013; accepted April 4, 2013.

Author contributions: S.v.S. and U.N. designed research; S.v.S. performed research; S.v.S. and U.N. analyzed data; S.v.S. and U.N. wrote the paper.

This work was supported by the Max-Planck-Society. We thank Sebastian Werner, Hweeling Lee, Johannes Tünnerhoff, and Ruth Adam for help.

The authors declare no competing financial interests.

Correspondence should be addressed to Sebastian von Saldern, Braystrasse 14, 81677 Munich, Germany. E-mail: sebastian.von-saldern@gmx.de.

DOI:10.1523/JNEUROSCI.3020-12.2013

Copyright © 2013 the authors 0270-6474/13/338841-09\$15.00/0

conditions, nor can they provide insights into the operational modes governing multisensory integration. Since the seminal work by Stein et al. (Meredith and Stein, 1986; Wallace et al., 1996; Stanford et al., 2005), numerous studies have shown that sensory inputs can be combined in a superadditive, additive, or subadditive fashion. According to the principle of inverse effectiveness, the operational integration mode depends on stimulus efficacy, with superadditive responses being the most pronounced when the unisensory stimuli are the least effective. Most recently, it has been suggested that the principle of inverse effectiveness may result from divisive normalization in which the activity of each individual neuron is divided by the net activity of all neurons to generate the final response (Ohshiro et al., 2011).

To address these questions, we presented participants with auditory, visual, and audiovisual motion stimuli in a motion discrimination task at two different levels of signal reliability. First, we tested for response nonlinearities (i.e., superadditive or subadditive interactions) at different reliability levels to identify audiovisual integration in sensory, association, and response-related areas. Second, we investigated whether these operational integration modes depend on signal efficacy as expected under the principle of inverse effectiveness.

Materials and Methods

Subjects. After giving informed consent, 20 healthy right-handed German native speakers (6 females; mean age: 26.45 ± 3.35 years) participated in this fMRI study. All subjects had normal or corrected to normal vision and reported normal hearing. The study was approved by the joint human research review committee of the Max-Planck-Society and the University of Tübingen.

Auditory stimulus. The auditory stimulus was a white noise sound of 500 ms duration, low-pass filtered with a fifth-order Butterworth filter at a cutoff frequency of 2 kHz. In contrast to previous studies (Alais and Burr, 2004), continuous auditory motion was created by providing two binaural (i.e., amplitude and latency) cues to mimic natural auditory motion signals more closely. Amplitude cues were provided by fading the sound amplitude between the left and right ear channels from 100% to 50% over the course of one trial. Latency cues were provided by presenting the sounds with the appropriate interaural time difference to the two ears. This led to the impression of a sound source moving from -8.5° to $+8.5^\circ$ along the azimuth. We did not include monoaural spectrotemporal filtering cues mimicking the filtering of the pinna that are primarily relevant for discriminating motion along the vertical dimension, because in prior psychophysics pilot studies, continuous motion stimuli with only binaural cues proved to be more powerful than apparent motion stimuli with additional monoaural spectrotemporal filtering cues.

Finally, we manipulated the reliability of the directional motion signal by jittering different amounts of auditory frames. To create auditory noise, we completely randomized the order of auditory samples.

Visual stimulus. The visual stimulus was a leftwards or rightwards moving disc (2.83° visual angle) defined by “shape from motion.” For this, a random noise rectangle ($16.94^\circ \times 3.71^\circ$) was generated for each frame (i.e., at a refresh rate of 60 Hz) by sampling the intensity values for each pixel uniformly within a range from 0 (black, luminance 0.79 cd/m^2) to 1 (white, luminance 128 cd/m^2). Within this rectangle, the disc was defined by intensity values that were randomly sampled at the beginning of the trial and then held constant throughout the course of the 500 ms trials. Directional leftwards or rightwards motion was created by displacing these constant intensity values from frame to frame by 1000 pixels/s (i.e., at a velocity of $33.88^\circ/\text{s}$) along the azimuth in a spatially coherent fashion. Because the distribution of intensity values was identical inside and outside of the disc in each frame, the disc merged with the background in every single frame. As a shape from motion stimulus, the disc could only be segmented from the background by integrating the intensity values over several frames. In that way, the perception of directional motion and the perception of the disc were tightly intertwined. In contrast to conventional random dot kinematograms, this shape from

motion disc enables strong colocalization cues that are thought to be essential for audiovisual integration (Meyer et al., 2005). Furthermore, this structure from motion stimulus to some extent mimicked our natural environment, where humans need to discriminate the motion direction of visual stimuli that are camouflaged by dynamic noise.

The reliability of directional visual motion was manipulated by changing the fraction of pixels within the disc that were held at a constant luminance value throughout the course of the trial. To create visual noise, this fraction was set to zero; that is, all pixels were assigned a new random luminance value sampled independently for each frame.

The random noise rectangle was presented on a gray background (luminance 32.6 cd/m^2). A red fixation cross (0.189° visual angle) was presented in the center of the screen throughout the entire experiment. Luminance values are based on measurements in the psychophysical setup used for training. In this setup, the Minolta CS-100 chromameter, which is MRI incompatible, was used so that no measurements could be performed inside the scanner.

Experimental design and procedures. In a motion discrimination task, subjects were presented with audiovisual inputs that contained a directional motion signal in the auditory, visual, both, or neither of the two modalities. Therefore this experiment focused specifically on the integration of directional motion information while controlling for activations pertaining to integration of audiovisual signals per se. In the audiovisual conditions, auditory and visual motion directions were always congruent.

Specifically, the $2 \times 2 \times 2$ factorial design manipulated: (1) auditory motion information (signal vs noise), (2) visual motion information (signal vs noise), and (3) the reliability of motion information (intact vs degraded). Because the auditory noise + visual noise condition was identical for the two levels of signal reliability, this factorial design resulted in seven conditions: auditory intact + visual intact ($A_i V_i$), auditory intact + visual noise ($A_i V_n$), auditory noise + visual intact ($A_n V_i$), auditory noise + visual noise ($A_n V_n$), auditory degraded + visual degraded ($A_d V_d$), auditory degraded + visual noise ($A_d V_n$), and auditory noise + visual degraded ($A_n V_d$). This design was used to maximize the number of trials that enter into the critical inverse effectiveness contrast, as follows: $A_i V_i - (A_i V_n + A_n V_i) \neq A_d V_d - (A_d V_n + A_n V_d)$ (see below).

The intact, degraded, and noise stimulus components were generated by manipulating the signal's reliability using shuffling and randomization procedures in time and space (see visual and auditory stimuli). Therefore, the noise conditions were generated from the signal conditions by abolishing the directional motion signal (or approximately by setting the variance of the motion information to infinity). For brevity, we refer to conditions that contain directional motion signal only in one sensory modality as unisensory motion conditions (even though the sensory inputs were always given in both auditory and visual modalities).

In a two-alternative forced choice task, subjects discriminated the direction of the motion signal (right vs left), indicating their response as fast and accurately as possible via a two-choice key press. The duration of the stimulus was 500 ms, followed by a 1900 ms fixation period, yielding a stimulus onset asynchrony of 2400 ms. In total, there were 560 trials arranged in four sessions with 20 trials per condition and session. Each session included 10 left and 10 right motion trials per condition. Trials were arranged in blocks of 14 stimuli interleaved with 7 s fixation. Whereas the motion direction of each trial was fully randomized, the events of the seven activation conditions were presented in a pseudorandomized fashion to maximize design efficiency.

Adaptive staircase procedures. After initial training sessions outside of the scanner, the signal reliability of the degraded conditions was determined before the main experiment inside the scanner (i.e., with scanner noise present) using the adaptive staircase procedure QUEST with a cumulative Gaussian as the underlying psychometric function (Watson and Pelli, 1983). For each subject, the amount of shuffling for auditory stimuli and the fraction of coherently moving luminance values for visual stimuli were adjusted individually to obtain 80% accuracy in the degraded unisensory conditions. Therefore, four unisensory sessions (80 trials) with separate, interleaved QUEST procedures were performed for unisensory auditory and visual trials. The threshold level was estimated

for each participant by averaging of at least three QUEST procedures. The intensive QUEST procedure inside the scanner allowed us to adjust the motion signal reliability individually in audition and vision to 80% accuracy. We thereby maximized the chances to obtain an audiovisual motion benefit, because under the assumptions of maximum likelihood integration, the reduction in variance for the audiovisual relative to the unisensory conditions is maximal, when the variance (and hence the d' in the motion discrimination task) of the auditory and visual conditions are equated.

To ensure that training and exposure were equated for unisensory and audiovisual conditions, subjects also participated in four audiovisual sessions with predefined motion intensities.

Stimulus presentation. Visual and auditory stimuli were presented using Cogent (John Romaya, Vision Laboratory, University College London; <http://www.vislab.ucl.ac.uk/>), running under MATLAB 7.0 (MathWorks) on a Windows PC. Auditory stimuli were presented at ~80 dB SPL using MR-compatible headphones (MR Confon). Visual stimuli were back-projected onto a Plexiglas screen using an LCD projector (JVC) visible to the subject through a mirror mounted on the MR head coil. Subjects performed a behavioral task using an MR-compatible custom-built button device connected to the stimulus computer.

Eye tracking. Eye movement data were collected inside the scanner for six subjects during all four sessions of the experiment using an ASL EyeTrack 6 Long Range Optics System (Applied Science Laboratories) and analyzed using routines custom written in MATLAB. The sampling frequency of the eye tracker was 60 Hz; the eye-recording system was calibrated using a 9-point calibration screen with eccentricities between 2.64° and 4.5°. Eye position data were *post hoc* corrected for constant offset and blinks and converted to velocity. One subject was excluded from the analysis because the trigger signals were not recorded. For each trial, we computed: (1) the distance (visual degree) from fixation mean, (2) the number of saccades per second (velocity threshold > 120° per second), and (3) the number of pursuit eye movements. Pursuit eye movements were defined as eye movements that were consistent with the stimulus in direction and velocity (33.88°/s; allowing for a 10% tolerance in eye velocity) for a duration of at least 50 ms. In addition, we calculated the average number of trials per condition contaminated by eye movements (including saccades and pursuit eye movements).

fMRI data acquisition. A 3T Siemens Magnetom Trio System was used to acquire both a T1-weighted three-dimensional high-resolution anatomical image (TR = 2300 ms, TE = 2.98 ms, TI = 1100 ms, flip angle = 9°, FOV = 256 × 240 × 176 mm, image matrix = 256 × 240 × 176, isotropic spatial resolution 1 mm) and T2*-weighted axial echoplanar images with BOLD contrast [gradient echo, TR = 3080 ms, TE = 40 ms, flip angle = 90°, FOV = 192 × 192 mm, image matrix 64 × 64, 38 axial slices acquired in ascending order, covering only the cerebrum without the cerebellum, voxel size = 3.0 mm × 3.0 mm × (2.5 mm + 0.5 mm) interslice gap]. There were four sessions with a total of 138 volume images per session. The first three volumes in each session were discarded to allow for T1-equilibration effects.

fMRI data analysis. The data were analyzed with statistical parametric mapping using SPM5 software from the Wellcome Trust Centre of NeuroImaging, London (<http://www.fil.ion.ucl.ac.uk/spm>; Friston et al., 1994). Scans from each subject were realigned using the first as a reference, unwarped, spatially normalized into MNI standard space (Evans et al., 1992), resampled to 2 × 2 × 2 mm³ voxels, and spatially smoothed with a Gaussian kernel of 8 mm FWHM. The time series of all voxels were high-pass filtered to 1/128 Hz. The fMRI experiment was modeled in an event-related fashion with regressors entered into the design matrix after convolving each trial with a canonical hemodynamic response function and its first temporal derivative. In addition to modeling the 7 conditions in our 2 × 2 × 2 factorial design, the statistical model included the realignment parameters as nuisance covariates (to account for residual motion artifacts). Condition-specific effects for each subject were estimated according to the general linear model and passed to a second-level ANOVA as contrasts.

At the second between-subjects level, we tested for the following effects:

(1) Main effects of auditory and visual motion:

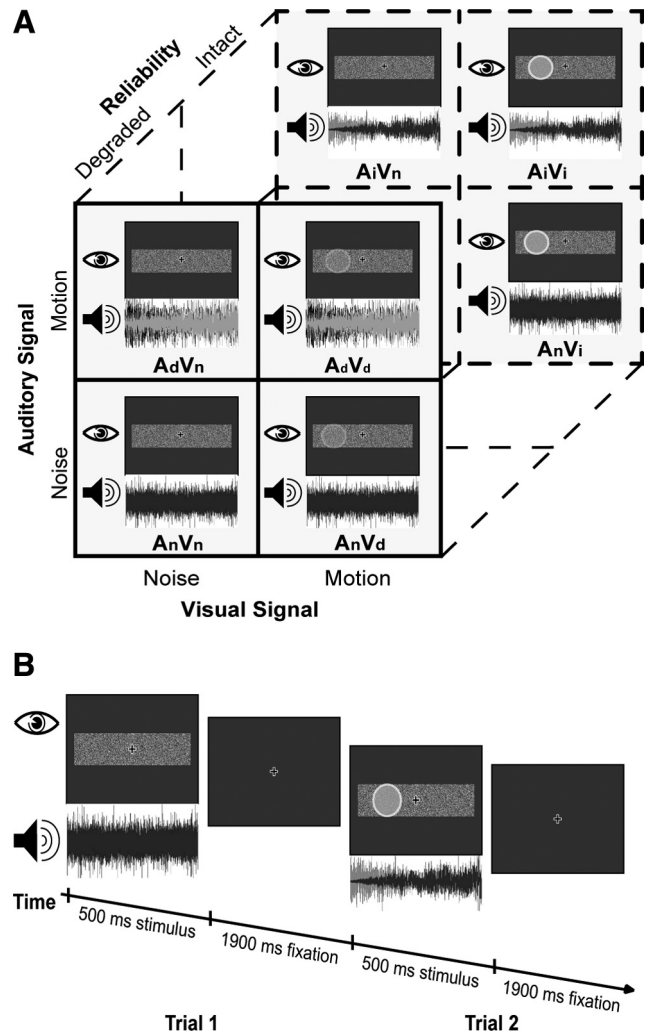


Figure 1. Experimental design and procedure. **A**, The 2 × 2 × 2 factorial design with the following factors: auditory signal (the presence of motion information versus noise); visual signal (the presence of motion information versus noise); and reliability (degraded stimuli versus intact stimuli). Example stimuli are presented as images of the visual stimulus and the corresponding sound waveforms. The visual moving disc is shaded light gray for illustrational purposes only. In the experiment, the disc as a structure from motion stimulus was not discernible from the background in a single frame. **B**, Timeline for two example trials. Subjects are presented with audiovisual stimuli that contain a motion signal in the auditory, the visual, both, or none of the two modalities. In a two-alternative forced choice task, they discriminate the motion direction for each 500 ms stimulus (left vs right). For illustrational purposes only, the moving visual disc is shaded light gray.

We identified areas involved in directional auditory motion processing by comparing all auditory motion with auditory noise conditions: $A_iV_i + A_iV_n + A_dV_d + A_dV_n > A_nV_i + A_nV_d + 2 * A_nV_n$. To ensure a valid and balanced statistical comparison, the noise condition needs to enter this contrast twice, because it is identical for both levels of signal reliability (Fig. 1). Likewise, we identified areas involved in directional visual motion processing by comparing all visual motion with visual noise conditions: $A_iV_i + A_nV_i + A_dV_d + A_nV_d > A_iV_n + A_dV_n + 2 * A_nV_n$.

(2) Subadditive and superadditive interactions:

We identified areas involved in audiovisual integration by testing for subadditive and superadditive interactions separately for each level of reliability (e.g., $A_iV_i + A_nV_n < A_iV_n + A_nV_i$). Based on our previous studies of multisensory object categorization, we expected primarily subadditive interactions for intact stimuli. However, for completeness, we also tested for superadditive interactions.

(3) Inverse effectiveness:

Table 1. Behavioral results

	A_dV_n	A_nV_d	A_dV_d	A_iV_n	A_nV_i	A_iV_i	A_nV_n
Accuracy (% correct)							
Mean	80.3	77.2	92.4	91.5	94.6	97.1	50.8
SEM	2.0	2.4	1.3	2.3	2.4	1.1	1.3
Response time (ms)							
Mean	884.3	824.3	801.7	831.0	720.6	715.2	886.5
SEM	27.0	32.5	28.7	24.0	25.8	25.2	27.6
d' (arbitrary units)							
Mean	1.9	1.7	3.1	3.1	3.4	3.6	0.1
SEM	0.2	0.2	0.2	0.2	0.2	0.1	0.1

Accuracy, response times, and d' values are shown for the motion discrimination task. Values represent across-subjects means and SEM.

Finally, we investigated whether the operational profile of multisensory integration depends on signal reliability. Consistent with the principle of inverse effectiveness, we expected a subadditive or even suppressive integration profile for intact stimuli to turn into additive or even superadditive interaction profiles for degraded stimuli. Therefore, we investigated whether subadditivity was increased for intact relative to degraded stimuli as follows: $\text{subadditivity}_{\text{intact}} > \text{subadditivity}_{\text{degraded}}$: $A_iV_n + A_nV_i - A_iV_i > A_dV_n + A_nV_d - A_dV_d$.

In a second step, we further characterized the subadditive and superadditive interactions as audiovisual response enhancement and suppression by comparing the audiovisual response with the maximal unisensory response. For example, subadditive interactions can be associated both with response enhancement and response suppression of the audiovisual relative to the maximal unisensory response. Again, based on the principle of inverse effectiveness, we would expect that response suppression for intact stimuli turns into response enhancement for degraded stimuli. Please note that this additional analysis serves only as a depth characterization of our data, because the statistical comparisons pertaining to the interaction and response enhancements/suppressions are not orthogonal to each other.

Search volume constraints. We tested for each effect within the neural systems (within the cerebrum) engaged in audiovisual processing; that is, corrected for multiple comparisons within all areas that were activated for all conditions $>$ fixation at $p < 0.001$ uncorrected ($VOI_{\text{All}>\text{Fix}} = 6486$ voxels). In addition, we used two more specific search volume constraints based on our a priori hypotheses that audiovisual integration of motion information may emerge in auditory and visual areas involved in motion processing.

Given the prominent role of hMT+/ $V5+$ in visual motion processing, the first search volume was defined bilaterally as the two clusters centered on the peak voxels at [48, -64, 0] and [-46, -72, 2] in the main effect of visual motion thresholded at $p < 0.001$, masked with all $>$ fixation at $p < 0.001$ uncorrected ($VOI_{\text{visual}} = 162$ voxels). (Based on cytoarchitectonic probability mapping, these two peak voxels were assigned to hMT with a probability of 30% and 40%; Eickhoff et al., 2005).

Given the role of posterior superior temporal gyrus/planum temporale in auditory motion processing (Baumgart et al., 1999; Pavani et al., 2002; Warren et al., 2002), the second search volume was defined as the two clusters centered on the peak voxels at [68, -30, 12] and [-52, -26, 8] in the main effect of auditory motion thresholded at $p < 0.001$, masked with all $>$ fixation at $p < 0.001$ uncorrected ($VOI_{\text{audio}} = 497$ voxels). All search volumes were created using the MarsBaR (<http://marsbar.sourceforge.net/>) toolbox (Brett et al., 2002).

Please note that the all conditions $>$ fixation or the auditory and visual motion contrasts used to define these search volumes are orthogonal to the interaction contrasts, thereby enabling valid and unbiased statistical inferences (Friston et al., 2006; Kriegeskorte et al., 2009). Unless otherwise stated, we report activations at $p < 0.05$ at the voxel level corrected for multiple comparisons within the respective search volume.

Results

Behavioral results

During the functional imaging study, subjects discriminated between rightward and leftward motion (Table 1). A two-way

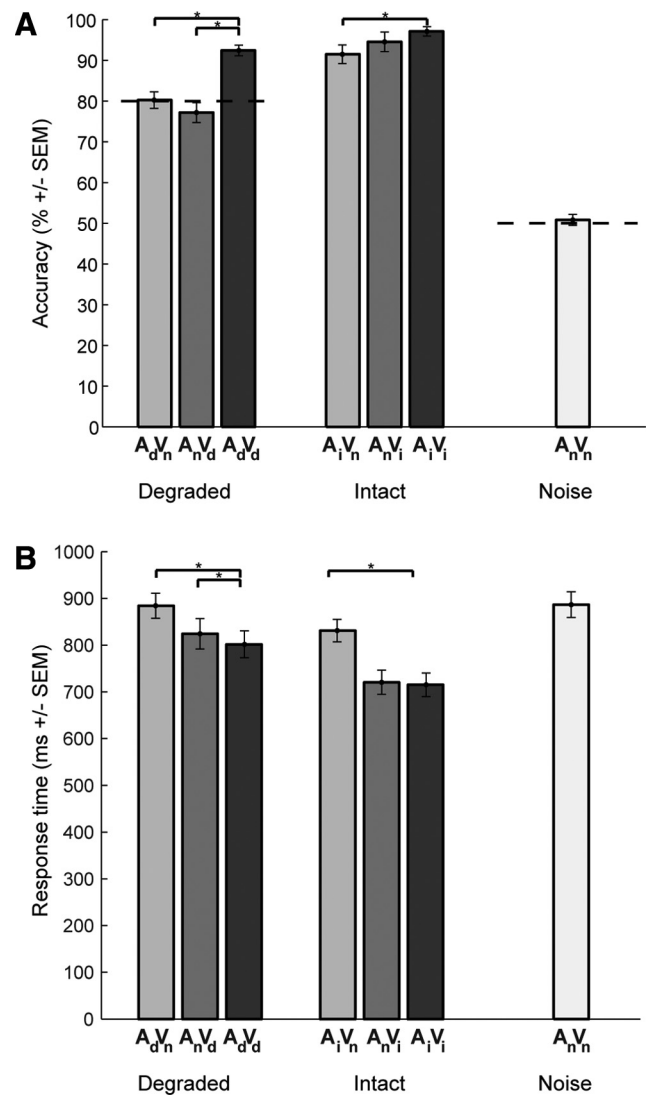


Figure 2. Behavioral results. Accuracy and response times for motion discrimination are presented as bar plots for degraded, intact, and noise conditions. **A**, Discrimination accuracy as across-subjects mean (% correct) \pm SEM. Dotted lines indicate 80% threshold in degraded conditions and 50% chance level in noise condition. **B**, Response times as across-subjects mean (ms) \pm SEM.

repeated-measures (RM) ANOVA with the factors modality (auditory, visual, audiovisual) and sensory reliability (degraded, intact) for performance accuracy identified significant main effects of modality ($F_{(1,251, 23,762)} = 24.078$; $p < 0.001$) and sensory reliability ($F_{(1,19)} = 30.363$; $p < 0.001$, Greenhouse-Geisser corrected). There was also a significant interaction effect between both factors ($F_{(1,533, 29,121)} = 11.834$; $p < 0.001$, Greenhouse-Geisser corrected).

For response times, the two-way RM ANOVA also showed significant main effects of modality ($F_{(1,225, 23,272)} = 48.619$; $p < 0.001$) and sensory reliability ($F_{(1,19)} = 27.710$; $p < 0.001$), as well as an interaction between them ($F_{(1,671, 31,758)} = 6.489$; $p < 0.001$, Greenhouse-Geisser corrected).

To investigate audiovisual benefits, we compared (paired t test) performance accuracy and response times of the audiovisual relative to the auditory and visual conditions (Fig. 2). A significant increase in performance accuracy was observed for the degraded audiovisual condition relative to both the auditory ($t_{19} = 8.5709$; $p < 0.001$) and the visual condition ($t_{19} = 7.7519$;

Table 2. Eye movement results

	A_dV_n	A_nV_d	A_dV_d	A_iV_n	A_nV_i	A_iV_i	A_nV_n
Mean distance from fixation mean (visual degree)							
Mean	0.57	0.61	0.63	0.63	0.61	0.63	0.65
SEM	0.07	0.07	0.08	0.09	0.09	0.1	0.09
Number of saccades per second							
Mean	0.29	0.24	0.43	0.44	0.27	0.3	0.43
SEM	0.11	0.06	0.15	0.13	0.06	0.08	0.1
Number of trials contaminated with eye movement							
Mean	1.35	1.5	1.7	1.9	1.1	1.45	1.95
SEM	0.39	0.39	0.46	0.49	0.26	0.46	0.41

Mean distance from fixation mean (visual degree), number of saccades per second, and number of trials contaminated with eye movement (saccades or smooth pursuit movements) are shown for the motion discrimination task. Values represent across-subjects means and SEM.

Table 3. fMRI results: auditory and visual motion processing

Regions	MNI coordinates			z-score	p_{FWE}^a
	x	y	z		
Auditory motion ($A_iV_i + A_iV_n + A_dV_d + A_dV_n > A_nV_i + A_nV_d + 2 * A_nV_n$)					
Right superior temporal gyrus/right planum temporale	68	−30	12	6.42	0.000
Left superior temporal gyrus/left planum temporale	−52	−26	8	5.04	0.000
Left putamen	−26	−2	−8	4.42	0.006
Visual motion ($A_iV_i + A_nV_i + A_dV_d + A_nV_d > A_iV_n + A_dV_n + 2 * A_nV_n$)					
Left middle temporal/occipital gyrus/V5+/hMT+	−46	−72	2	>7	0.000
Left middle occipital gyrus	−34	−88	0	6.67	0.000
Right middle temporal gyrus/V5+/hMT+	48	−64	0	>7	0.000
Right middle occipital gyrus	36	−84	4	6.90	0.000
Left putamen	−26	−2	−10	5.84	0.000
Right fusiform gyrus	28	−52	−20	4.39	0.007

^a p_{FWE} indicates *p*-value corrected at the voxel level for multiple comparisons within the following search volume: $VOI_{All>Fix}$ (6486 voxels).

$p < 0.001$). For the intact conditions, a significant increase could only be observed for the audiovisual relative to the auditory condition ($t_{19} = 3.7438$; $p < 0.01$). A similar pattern was found for response times, in which a significant decrease could be found for the degraded audiovisual relative to the auditory ($t_{19} = 7.0881$; $p < 0.001$) and the visual conditions ($t_{19} = 2.4726$; $p = 0.023$), whereas in the intact conditions, a significant decrease was only found for the audiovisual relative to the auditory condition ($t_{19} = 6.7725$; $p < 0.001$).

Eye movement results

Eye movements were characterized using four indices: (1) the distance (visual degree) from fixation mean, (2) the number of saccades per second (velocity threshold > 120°/s), (3) the number of pursuit eye movements, and (4) the number of trials contaminated by eye movements (including saccades and pursuits; Table 2).

Consistent with the behavioral data, all indices were entered into a two-way RM ANOVA with the factors modality (auditory, visual, audiovisual) and reliability (degraded, intact). [Please note that we used a 2 × 3 factorial ANOVA rather than a 2 × 2 × 2 factorial ANOVA reflecting our experimental design (Fig. 1) to focus the analysis on the conditions that entered into the inverse effectiveness contrast central to our fMRI analysis.]

For the deviation from the fixation mean, the analysis did not show main effects of modality ($F_{(1.726, 6.905)} = 1.400$; $p = 0.303$) or reliability ($F_{(1,4)} = 0.432$; $p = 0.547$) or an interaction between the two ($F_{(1.045, 4.181)} = 0.339$; $p = 0.600$; all results reported Greenhouse-Geisser corrected). Likewise, for number of saccades per second, this analysis identified no significant main effects of modality ($F_{(1.645, 6.579)} = 2.111$; $p = 0.196$) or reliability ($F_{(1,4)} = 0.769$; $p = 0.430$) or an interaction between the two ($F_{(1.202, 4.809)} = 3.165$; $p = 0.137$). Because we did not detect any pursuit

movements (see Materials and Methods for criteria to identify pursuits), no further analysis was performed for pursuit movements. Finally, for the trials contaminated with eye movement, the analysis did not reveal any significant main effects of modality ($F_{(1.897, 7.588)} = 0.791$; $p = 0.481$) or reliability ($F_{(1,4)} = 0.082$; $p = 0.789$) or an interaction between the two ($F_{(1.544, 6.176)} = 3.154$; $p = 0.119$).

These findings indicated that subjects' eye movements did not differ across conditions, rendering it unlikely that eye movements account for the activation differences observed in the fMRI data.

fMRI results

The fMRI analysis was performed in three steps. First, we identified regions that were involved in processing of directional auditory or visual motion signals. Second, we identified regions involved in audiovisual integration of directional motion signals by testing for superadditive and subadditive interactions separately for intact and degraded stimuli. Finally, we directly compared the interactions for degraded and intact stimuli to investigate whether the integration profile followed the principle of inverse effectiveness.

Auditory and visual motion processing

As expected, directional visual motion signals increased activations bilaterally in the middle temporal gyri (hMT/V5+, the classical visual motion areas; Table 3). Further activations were observed in the left putamen, the middle occipital gyri bilaterally, and in the right fusiform gyrus. Conversely, directional auditory motion relative to auditory noise increased activations in superior temporal gyri and plana temporalia bilaterally, as well as in the left putamen (Fig. 3).

Audiovisual interactions

For intact stimuli, we observed significant subadditive interactions in the left putamen, the right superior temporal gyrus/planum temporale (auditory motion area), and the left middle occipital/temporal gyrus (visual motion area) corresponding to left hMT+/V5+. At an uncorrected threshold of significance, we also observed subadditive interactions in the left superior temporal gyrus/planum temporale and the right hMT+/V5+, suggesting that motion information is integrated along the auditory and visual motion processing streams bilaterally (Table 4). We did not observe any superadditive interactions for degraded or intact stimuli. This replicates and extends our previous findings in object categorization, where we did not observe any reliable superadditive interactions at the random-effects group level (Werner and Noppeney, 2010a,b).

Inverse effectiveness

To investigate whether the operational mode of audiovisual motion integration depends on stimulus efficacy, as predicted by the principle of inverse effectiveness, we compared the interactions for intact and degraded stimuli directly. Indeed, consistent with the principle of inverse effectiveness, the audiovisual interactions in the left putamen, left superior temporal gyrus/planum temporale, and left middle temporal gyrus were more subadditive for intact than degraded stimuli. In fact, the integration profile was suppressive for intact stimuli (i.e., the audiovisual response was less than the maximal unisensory response); however, it turned into response enhancement for degraded stimuli, where the audiovisual response was greater than the maximal unisensory response (Fig. 4).

Again, at an uncorrected level of significance, the same response profile was observed in the homolog regions of the contralateral hemisphere: the right superior temporal gyrus/planum temporale and the right middle temporal gyrus.

Discussion

Our results demonstrate that sensory areas, together with the putamen, integrate auditory and visual motion information into behavioral benefits for motion discrimination. Consistent with the principle of inverse effectiveness, the operational mode of integration in these areas depended on the sensory reliability and stimulus efficacy, with the subadditive integration profile being most pronounced for intact stimuli.

At the behavioral level, subjects benefited significantly from audiovisual integration primarily for degraded motion signals while obtaining near ceiling performance for unisensory signals when these were reliable and intact. In contrast to previous studies, our experimental design manipulated the presence versus absence of directional motion information in the auditory and visual modalities rather than of the signal per se. Therefore, in unisensory trials, directional motion information was limited to one sensory modality, with the other modality presenting only noise. The interfering and distracting effects of the noise in the unisensory conditions most likely explains the pronounced increase in motion discrimination accuracy and d' for degraded audiovisual relative to unisensory motion signals, which exceeded the purely statistical audiovisual benefits (e.g., maximum

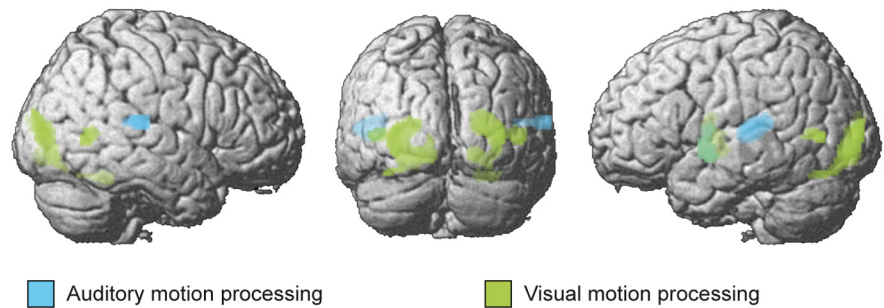


Figure 3. Auditory and visual motion processing. Activations for auditory and visual motion processing rendered on a whole-brain template. Blue indicates auditory motion ($A_i V_i + A_i V_n + A_d V_d + A_d V_n > A_n V_i + A_n V_d + 2 * A_n V_n$); green, visual motion ($A_i V_i + A_n V_i + A_d V_d + A_n V_d > A_i V_n + A_d V_n + 2 * A_n V_n$). Height threshold: $p < 0.001$ and masked with all $>$ fixation at $p < 0.001$ uncorrected for illustrational purposes.

likelihood integration) that were observed in previous studies (Meyer and Wuerger, 2001; Wuerger et al., 2003; Alais and Burr, 2004).

At the neural level, audiovisual interactions were observed in audiovisual areas and the putamen. More specifically, subadditive interactions emerged in visual motion area hMT+/V5+ and the planum temporale/posterior superior temporal gyrus, which are thought to be involved in auditory motion processing (Baumgart et al., 1999; Pavani et al., 2002; Warren et al., 2002). A multisensory role has been proposed previously for these classical unisensory motion areas using paradigms of visual motion capture. However, visual motion capture was associated with activations in the parietal cortex followed by a concurrent activation increase in hMT+/V5+ and activation decrease in planum temporale (Alink et al., 2008). The seesaw relationship between auditory and visual activations suggests that motion capture may rely on attentional shifts from the auditory to the visual modality. In contrast, our study revealed subadditive integration profiles concurrently in both auditory and visual areas in the absence of parietal activations, suggesting integration mechanisms at the level of sensory motion processing. Indeed, recent studies have accumulated evidence for low-level multisensory integration in putatively unisensory areas (Schroeder and Foxe, 2002; van Atteveldt et al., 2004; Macaluso and Driver, 2005; Murray et al., 2005; Schroeder and Foxe, 2005; Bonath et al., 2007; Kayser et al., 2007; Lakatos et al., 2007; Lewis and Noppeney, 2010).

The absence of audiovisual interactions in parietal cortices may be surprising given previous studies showing parietal activations for auditory, visual, and tactile motion relative to fixation (Lewis et al., 2000; Bremmer et al., 2001). However, in contrast to previous research, we manipulated selectively the presence/absence of directional motion information rather than auditory/visual inputs per se. Furthermore, we presented noise trials and motion trials at different levels of reliability in a randomized fashion. Subjects were instructed to discriminate left versus right motion consistently for all trials, even for the pure noise trials that did not contain any directional motion signal. Under these circumstances, subjects are known to integrate random evidence for a particular perceptual interpretation into a superstitious perception (Gosselin and Schyns, 2003). Indeed, participants in our study also treated noise trials like near-threshold motion signals, as indicated by comparable response times. Therefore, our design equated the directional motion and noise conditions far more tightly with respect to sensory inputs, attentional demands, and cognitive set. It thereby minimized the factors that drove parietal activations in previous studies that compared auditory, visual, or tactile motion rela-

Table 4. fMRI results: audiovisual interactions and inverse effectiveness

Regions	MNI coordinates			z-score	p _{FWE}
	x	y	z		
Subadditivity for intact conditions ($A_iV_i + A_nV_n < A_iV_n + A_nV_i$)					
Left putamen	-28	2	-2	5.08	0.000 ^a
Right superior temporal gyrus/right planum temporale	68	-30	12	3.27	0.029 ^b
Left superior temporal gyrus/left planum temporale /left rolandic operculum	-38	-36	16	2.94	0.072 ^b
Left middle occipital/temporal gyrus/V5+ /hMT+	-36	-68	6	2.77	0.050 ^c
Right middle temporal gyrus/V5+ /hMT+	48	-60	0	2.40	0.118 ^c
Inverse effectiveness (subadditivity _{intact} > subadditivity _{degraded})					
Left putamen	-26	2	0	4.28	0.011 ^a
Left superior temporal gyrus/Left planum temporale ^d	-36	-40	14	4.08	0.023 ^a
	-38	-36	18	3.64	0.009 ^b
Right superior temporal gyrus/Right planum temporale	66	-32	12	2.86	0.087 ^b
Left middle temporal gyrus/V5+ /hMT+	-36	-68	4	2.99	0.028 ^c
Right middle temporal gyrus/V5+ /hMT+	42	-58	4	2.73	0.056 ^c

p_{FWE} indicates corrected at the voxel level for multiple comparisons within the following search volumes: ^aVOI_{All>Fix} (6486 voxels); ^bVOI_{audio} (497 voxels); ^cVOI_{visual} (162 voxels); ^dCluster extending into white matter.

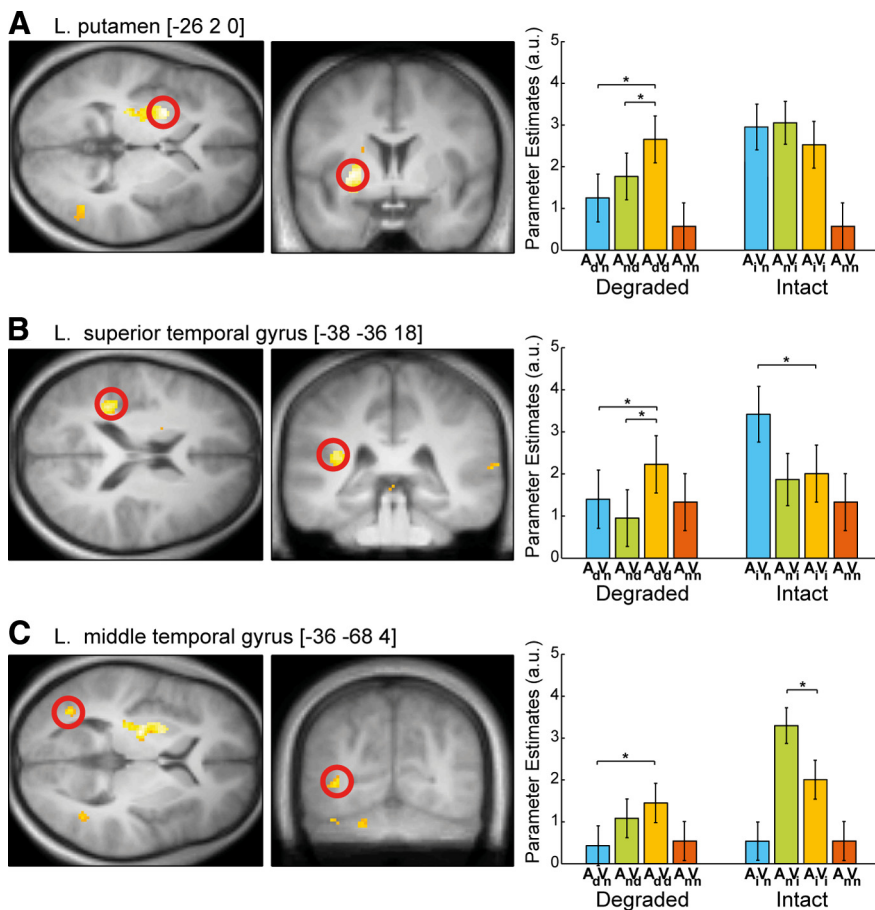


Figure 4. Inverse effectiveness. Left: Activations pertaining to inverse effectiveness (subadditivity_{intact} > subadditivity_{degraded}) on transversal and coronal slices of a mean structural image created by averaging the subjects' normalized structural images: left putamen (A), left superior temporal gyrus (B), and left middle temporal gyrus (C). Height threshold: $p < 0.01$ uncorrected, masked with all > fixation at $p < 0.01$ uncorrected for illustrational purposes. Peak voxels (i.e., the given coordinate locations) are marked with a red circle. Parameter estimates (mean \pm SEM) for degraded (A_dV_n/A_nV_d) and intact unisensory motion (A_iV_n/A_nV_i), audiovisual motion (A_iV_d/A_nV_d), and noise (A_nV_n) (plotted twice for illustrational purposes) conditions at the given coordinate locations are shown. The bar graphs represent the size of the effect in nondimensional units (arbitrary units [a.u.]) corresponding to percentage whole-brain mean.

tive to fixation (Bremmer et al., 2001). Indeed, even auditory motion and visual motion did not increase activations in parietal cortices relative to audiovisual noise trials (Fig. 3).

In addition to sensory cortices, subadditive interactions for intact motion signals were observed in the putamen bilaterally.

Whereas the basal ganglia generally play a key role in action selection (Middleton and Strick, 2000), the putamen is particularly involved in mapping sensory inputs onto overlearned responses (Alexander et al., 1986; Lehericy et al., 2004). Therefore, putaminal activations emerge primarily when the sensory motor mapping has become automatic during late stages of sensory motor learning (Amiez et al., 2012). Indeed, motion discrimination also relies on a simple and overlearned sensory motor mapping from left/right motion onto left/right responses. Therefore, our study demonstrates that the putamen integrates auditory and visual inputs into rapid responses in highly automatic tasks. In contrast, previous findings showed subadditive and even suppressive interactions in the prefrontal cortex during explicit audiovisual object categorization (Werner and Noppeney, 2010b) or perception of communication signals (Sugihara et al., 2006). These results suggest that auditory and visual signals may be integrated in different circuitries underlying response selection depending on stimulus complexity, task, and automaticity of response selection. The prefrontal cortex may integrate audiovisual signals during recognition of complex objects or communication signals that rely on an arbitrary mapping of stimulus onto category or response. In contrast, the putamen may be engaged when sensory signals are integrated at the response level during highly automatized tasks such as motion discrimination to enable effective and rapid interactions with our multisensory environment.

In terms of neural mechanisms, recent research has nuanced the role of superadditivity (Stein and Meredith, 1993) and emphasized the importance of additive and subadditive integration modes, especially for suprathreshold stimuli (Stanford et al., 2005; Stanford and Stein, 2007; Morgan et

al., 2008; Angelaki et al., 2009). Likewise, in our study, intact stimuli elicited subadditive interactions in hMT+/V5+, planum temporale, and the putamen in the absence of any superadditive interactions. The audiovisual interactions for intact motion signals were not only subadditive, but even suppressive: the response to bisensory stimuli was less than the response to the most effective unisensory stimulus. These suppressive interactions cannot easily be explained by saturation or other nonlinearities in the BOLD response (Buxton et al., 2004), but rather suggest sub-linear neural mechanisms of audiovisual integration and are consistent with previous fMRI results of object categorization and speech recognition in which only subadditive interactions at the level of the BOLD response were observed (Werner and Noppeney, 2010a,b; Lee and Noppeney, 2011).

In the putamen and sensory areas, the operational mode of multisensory integration depended on the reliability or effectiveness of the sensory motion information, as expected under the classical principle of inverse effectiveness (Meredith and Stein, 1983; Stein and Meredith, 1993; Stanford and Stein, 2007; Stein and Stanford, 2008). Therefore, even though the interactions were always subadditive, they turned from response suppression for intact stimuli into a significant response enhancement for degraded stimuli. Despite the fundamental role of inverse effectiveness in multisensory research (Stanford and Stein, 2007; Stein et al., 2009), its mechanism and functional role are still debated. Recent modeling work showed that inverse effectiveness may emerge naturally from divisive normalization as a simple mathematical operation (Ohshiro et al., 2011) in which the final response of a neuron is generated by dividing the activity of each neuron by the net activity of all neurons within a region.

From an alternative perspective, we may ask whether the principle of inverse effectiveness is behaviorally relevant. What is the functional role of response suppression and enhancement for subjects' motion discrimination performance? The audiovisual profile of response suppression for intact stimuli and response enhancement for near-threshold-degraded stimuli is reminiscent of the BOLD response profile induced by priming. In priming, repetition of clearly perceptible stimuli induces response suppression, whereas priming of near-threshold stimuli induces response enhancement. It has been suggested that response suppression reflects behavioral facilitation for perceptible stimuli, whereas response enhancement mediates the formation of a novel representation for stimuli that are not recognized without priming (Dolan et al., 1997; George et al., 1999; Henson, 2003). Likewise, in our multisensory context, audiovisual response suppression and enhancement may serve two distinct aims. For intact motion stimuli, in which each auditory and visual component signal can already individually sustain motion discrimination, response suppression may reflect facilitation of motion discrimination, as indicated by a nonsignificant decrease in response times for intact audiovisual relative to unisensory stimuli (the effect may have been attenuated because of observers waiting with their response to the end of the stimulus). In contrast, for degraded motion stimuli, audiovisual enhancement may mediate the significant increase in motion discrimination accuracy and possibly even the detection of an otherwise imperceptible circle. However, given the correlative nature of fMRI studies, this account remains to be confirmed. In particular, we cannot determine whether the absence of a multisensory behavioral benefit emerges necessarily from suppressive neural response profiles or if it is a side effect of ceiling performance in the unisensory conditions.

Finally, attentional mechanisms may have contributed to some extent to this "inverse effectiveness profile" because intersensory attentional shifts may depend on sensory reliability. Therefore, subjects may have shifted their attention to audition or vision more intensively when the particular modality contains the signal rather than the noise during the intact compared with the degraded conditions. These intersensory attentional shifts may then induce an amplification of processing and activations in the relevant sensory cortices (e.g., MT+) for the intact unisensory signal conditions (e.g., $A_i V_n$, $A_n V_i$) and thereby contribute to the stronger subadditive interactions for the intact relative to the degraded conditions. However, for this intersensory attentional account, it is difficult to explain the audiovisual enhancement during the degraded conditions, suggesting that attentional mechanisms may be a contributing factor but not the central mechanism underlying inverse effectiveness.

In conclusion, our study demonstrates audiovisual interactions in putatively unisensory areas such as the hMT+/V5+ and planum temporale, as well as the putamen, suggesting audiovisual integration processes at the sensory and response selection levels. In all of these regions, the operational profile of audiovisual integration followed the principle of inverse effectiveness in which audiovisual response suppression for intact stimuli turned into response enhancement for degraded stimuli. This response profile parallels behavioral indices of audiovisual integration with significant audiovisual benefits selectively for the degraded conditions.

References

- Alais D, Burr D (2004) No direction-specific bimodal facilitation for audiovisual motion detection. *Brain Res Cogn Brain Res* 19:185–194. [CrossRef Medline](#)
- Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357–381. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Amiez C, Hadj-Bouziane F, Petrides M (2012) Response selection versus feedback analysis in conditional visuo-motor learning. *Neuroimage* 59:3723–3735. [CrossRef Medline](#)
- Angelaki DE, Gu Y, Deangelis GC (2009) Multisensory integration: psychophysics, neurophysiology, and computation. *Curr Opin Neurobiol*
- Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. *Cereb Cortex* 17:1433–1443. [CrossRef Medline](#)
- Baumgart F, Gaschler-Markefski B, Woldorff MG, Heinze HJ, Scheich H (1999) A movement-sensitive area in auditory cortex. *Nature* 400:724–726. [Medline](#)
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA (2007) Neural basis of the ventriloquist illusion. *Curr Biol* 17:1697–1703. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox. *Neuroimage* 16.
- Buxton RB, Uludağ K, Dubowitz DJ, Liu TT (2004) Modeling the hemodynamic response to brain activation. *Neuroimage* 23:S220–233. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335. [CrossRef Medline](#)

- Evans AC, Collins DL, Milner B (1992) An MRI-based stereotactic atlas from 250 young normal subjects. *Soc Neurosci Abstr* 18:179.4.
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1994) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2:189–210. [CrossRef](#)
- Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. *Neuroimage* 30:1077–1087. [CrossRef](#) [Medline](#)
- George N, Dolan RJ, Fink GR, Baylis GC, Russell C, Driver J (1999) Contrast polarity and face recognition in the human fusiform gyrus. *Nat Neurosci* 2:574–580. [CrossRef](#) [Medline](#)
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285. [CrossRef](#) [Medline](#)
- Gosselin F, Schyns PG (2003) Superstitious perceptions reveal properties of internal representations. *Psychological science* 14:505–509. [CrossRef](#) [Medline](#)
- Henson RN (2003) Neuroimaging studies of priming. *Progress in neurobiology* 70:53–81. [CrossRef](#) [Medline](#)
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional imaging reveals visual modulation of specific fields in auditory cortex. *J Neurosci* 27:1824–1835. [CrossRef](#) [Medline](#)
- Kim R, Peters MA, Shams L (2012) $0 + 1 > 1$: How adding noninformative sound improves performance on a visual task. *Psychological science* 23:6–12. [CrossRef](#) [Medline](#)
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535–540. [CrossRef](#) [Medline](#)
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279–292. [CrossRef](#) [Medline](#)
- Lee H, Noppeney U (2011) Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J Neurosci* 31:11338–11350. [CrossRef](#) [Medline](#)
- Léhéry S, Ducros M, Van de Moortele PF, Francois C, Thivard L, Poupon C, Swindale N, Ugurbil K, Kim DS (2004) Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Ann Neurol* 55:522–529. [CrossRef](#) [Medline](#)
- Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:873–888. [CrossRef](#) [Medline](#)
- Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J Neurosci* 30:12329–12339. [CrossRef](#) [Medline](#)
- Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in neurosciences* 28:264–271. [CrossRef](#) [Medline](#)
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389–391. [CrossRef](#) [Medline](#)
- Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56:640–662. [Medline](#)
- Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals. *Neuroreport* 12:2557–2560. [CrossRef](#) [Medline](#)
- Meyer GF, Wuerger SM, Röhrbein F, Zetsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Exp Brain Res* 166:538–547. [CrossRef](#) [Medline](#)
- Middleton FA, Strick PL (2000) Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain and cognition* 42:183–200. [CrossRef](#) [Medline](#)
- Morgan ML, Deangelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59:662–673. [CrossRef](#) [Medline](#)
- Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb Cortex* 15:963–974. [CrossRef](#) [Medline](#)
- Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. *Nat Neurosci* 14:775–782. [CrossRef](#) [Medline](#)
- 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. [CrossRef](#) [Medline](#)
- Sadaghiani S, Maier JX, Noppeney U (2009) Natural, metaphoric, and linguistic auditory direction signals have distinct influences on visual motion processing. *J Neurosci* 29:6490–6499. [CrossRef](#) [Medline](#)
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. *Curr Opin Neurobiol* 15:454–458. [CrossRef](#) [Medline](#)
- 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. [CrossRef](#) [Medline](#)
- Stanford TR, Stein BE (2007) Superadditivity in multisensory integration: putting the computation in context. *Neuroreport* 18:787–792. [CrossRef](#) [Medline](#)
- Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci* 25:6499–6508. [CrossRef](#) [Medline](#)
- Stein BE, Meredith A (1993) *The merging of the senses*. Cambridge, MA: MIT.
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266. [CrossRef](#) [Medline](#)
- Stein BE, Stanford TR, Ramachandran R, Perrault TJ Jr, Rowland BA (2009) Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp Brain Res* 198:113–126. [CrossRef](#) [Medline](#)
- Sugihara T, Diltz MD, Averbek BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138–11147. [CrossRef](#) [Medline](#)
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282. [CrossRef](#) [Medline](#)
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246–1266. [Medline](#)
- Warren JD, Zielinski BA, Green GG, Rauschecker JP, Griffiths TD (2002) Perception of sound-source motion by the human brain. *Neuron* 34:139–148. [CrossRef](#) [Medline](#)
- Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120. [CrossRef](#) [Medline](#)
- Werner S, Noppeney U (2010a) Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex* 20:1829–1842. [CrossRef](#) [Medline](#)
- Werner S, Noppeney U (2010b) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci* 30:2662–2675. [CrossRef](#) [Medline](#)
- Wuerger SM, Hofbauer M, Meyer GF (2003) The integration of auditory and visual motion signals at threshold. *Percept Psychophys* 65:1188–1196. [CrossRef](#) [Medline](#)