Contextual Control of Audiovisual Integration in Low-Level Sensory Cortices

Nienke M. van Atteveldt,^{1,2,3*} Bradley S. Peterson,³ and Charles E. Schroeder^{3,4}

¹Department of Cognitive Neuroscience, Maastricht University, Maastricht, The Netherlands

²Neuroimaging and Neuromodeling Group, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands

³Department of Psychiatry, New York State Psychiatric Institute, Columbia University, New York, New York

⁴Cognitive Neuroscience and Schizophrenia Program, Nathan S. Kline Institute, Orangeburg, New York

Abstract: Potential sources of multisensory influences on low-level sensory cortices include direct projections from sensory cortices of different modalities, as well as more indirect feedback inputs from higher order multisensory cortical regions. These multiple architectures may be functionally complementary, but the exact roles and inter-relationships of the circuits are unknown. Using a fully balanced context manipulation, we tested the hypotheses that: (1) feedforward and lateral pathways subserve speed functions, such as detecting peripheral stimuli. Multisensory integration effects in this context are predicted in peripheral fields of low-level sensory cortices. (2) Slower feedback pathways underpin accuracy functions, such as object discrimination. Integration effects in this context are predicted in higher-order association cortices and central/foveal fields of low-level sensory cortex. We used functional magnetic resonance imaging to compare the effects of central versus peripheral stimulation on audiovisual integration, while varying speed and accuracy requirements for behavioral responses. We found that interactions of task demands and stimulus eccentricity in low-level sensory cortices are more complex than would be predicted by a simple dichotomy such as our hypothesized peripheral/speed and foveal/accuracy functions. Additionally, our findings point to individual differences in integration that may be related to skills and strategy. Overall, our findings suggest that instead of using fixed, specialized pathways, the exact circuits and mechanisms that are used for low-level multisensory integration are much more flexible and contingent upon both individual and contextual factors than previously assumed. *Hum Brain Mapp* 35:2394–2411, 2014. © 2013 Wiley Periodicals, Inc.

Key words: fMRI; multisensory; visual cortex; auditory cortex; task demand; eccentricity

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Dutch Organization for Scientific Research (NWO, VENI); Contract grant number: 451-07-020 (to N.M.v.A.); Contract grant sponsor: European Community's Seventh Framework Programme [FP/2007-2013]; Contract grant number: 221187 (to N.M.v.A.); Contract grant sponsor: NIH; Contract grant number: R01 DC011490-11 (to C.E.S.).

*Correspondence to: N.M. van Atteveldt, Netherlands Institute for Neuroscience, Social Brain Lab, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands. E-mail: (N.vanAtteveldt@maastrichtuniversity.nl)

Received for publication 17 January 2013; Revised 7 May 2013; Accepted 15 May 2013.

DOI: 10.1002/hbm.22336

Published online 24 August 2013 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Combining the bits of information about an object or event that are conveyed through distinct sensory systems is an essential integrative operation the brain performs for a number of important reasons. Sometimes, for example, multisensory integration enables faster detection of a potentially dangerous object approaching in the periphery, such as a predator or a car. At other times, complementary information is combined to yield a richer internal representation and more accurate identification of complex sensory inputs, such as spoken language. Whether multisensory integration utilizes different brain circuits in such different contexts is unclear.

Substantial evidence from experiments in both human and nonhuman subjects suggests that multisensory interactions occur in low-level sensory areas, even at the level of primary visual and auditory cortices [Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007]. Three basic types of neuronal pathways have been suggested to support low-level multisensory integration [Schroeder et al., 2003]. In one case [e.g., Calvert, 2001], low-level interactions are mediated by feedback projections from higher-order multisensory cortices, such as the superior temporal cortex (STS) or intraparietal sulcus (IPS) [Smiley et al., 2007]. A second possibility is that these interactions are promoted by direct, lateral anatomical connections between low-level, even primary, sensory cortices [Bizley et al., 2007; Budinger et al., 2006; Cappe and Barone, 2005; Cappe et al., 2009a; Falchier et al., 2002, 2010; Rockland and Ojima, 2003]. A third alternative is that integration in low-level cortical areas is promoted by direct feedforward projections from nonspecific [Hackett et al., 2007; Lakatos et al., 2007] or higher order [Cappe et al., 2007] thalamic nuclei. The latter two views are supported by multisensory interactions at very short latencies [Cappe et al., 2010; Giard and Peronnet, 1999; Molholm et al., 2002; Raji et al., 2010; Schroeder et al., 2004].

Physiological evidence for low-level multisensory integration has accrued under a wide-range of experimental conditions. Collectively, these studies raise the possibility that different integration pathways are functionally complementary. For example, on the one hand, multisensory integration in low-level sensory cortices may use lateral (cortico-cortical) and feedforward (non-specific thalamic) projections to speed detection and localization of events, including potentially threatening ones coming from the periphery. Support for this idea comes from observations that: (1) lateral projections from auditory to visual cortex target peripheral rather than central representations of the visual field [Falchier et al., 2002, 2010], (2) cross-modal influences in humans appear stronger for peripheral than for central stimuli [Bolognini et al., 2010; Shams et al., 2001; Takeshima and Gyoba, 2013; Zhang and Chen, 2006] and have been linked specifically to fast behavioral reaction times (for auditory-tactile stimuli; [Sperdin et al., 2009], (3) compensatory plasticity after sensory loss,

putatively facilitated by unmasking of lateral cross-sensory connections [Budinger et al., 2006; Cappe and Barone, 2005; Lee et al., 2007], specifically enhances peripheral functions [Bavelier et al., 2006; Neville and Lawson, 1987; Röder et al., 1999]. On the other hand, cross-modal influences in low-level regions representing central visual/auditory space may result from feedback from higher-order regions to amplify stimulus representations useful when performing more complex tasks such as speech perception and object identification [Amedi et al., 2005; Van Atteveldt et al., 2004, 2009]. Evidence for feedback in this context comes from latencies lagging feedforward processing, and feedback laminar profiles [Schroeder et al., 2003]. It should be noted that potentially threatening centrally presented input has also been shown to elicit fast neural interactions in detection paradigms [Cappe et al., 2009b, 2012; Romei et al., 2009, in press]. This indicates the alternative possibility that speeded detection might be eccentricity-independent. In addition, the diverse findings of task influences [Fort et al., 2002; Van Atteveldt et al., 2007b; Werner and Noppeney, 2010] on low-level integration may suggest alternative specializations with regard to behavioral goal.

As a whole, these diverse findings support “functional specialization” hypotheses for different circuits for low-level multisensory integration. Such hypotheses have been articulated by several groups [Driver and Noesselt, 2008; Kayser and Logothetis, 2007; Schroeder et al., 2003], and have been examined to some extent for levels of object processing [Werner and Noppeney, 2010], but have not yet been addressed broadly and systematically in a single paradigm. Here, we test the predictions that: (1) multisensory integration in low-level auditory and visual cortices will be strongest for speeded detection of peripheral stimuli (i.e., in peripheral fields), and (2) integration in higher-order association cortices, and within central field representations of low-level cortices, will be strongest for tasks requiring accuracy in identification of centrally presented objects. We used functional Magnetic Resonance Imaging (fMRI) while presenting visual (V), auditory (A), or audiovisual (AV) objects either centrally or peripherally; and requiring participants to perform either a task that emphasized speeded detection or one that required accurate categorization. This fully balanced design allowed us to remain open to influences beyond the hypothesized accuracy/central versus speed/peripheral combinations, such as for example predicted by the early onset interactions during detection of central multisensory inputs and the different reported task effects. Furthermore, in contrast to some of the previous studies that presented peripheral stimuli at visual angles below 10° [Shams et al., 2001; Takeshima and Gyoba, 2013], we aimed to approach the eccentricities of peripheral fields reported by the macaque anatomical work [Falchier et al., 2002, 2010], to maximize the possibility to recruit the putative human equivalent of direct anatomical audiovisual projections.

MATERIALS AND METHODS

Participants

Twenty paid volunteers participated in this study. Two were excluded from analyses due to excessive head motion. The remaining 18 (seven females; average age 26.9, range 22–44) had normal vision or wore contact lenses. Adequate vision and audition were confirmed on the day of the experiment using a vision screener (Rosebaum pocket screener) and audiometer (Earscan 3 Manual audiometer, Micro Audiometrics Corp, Murphy, NC). All participants had vision within the normal range (distance equivalent 20/20 or 20/25) except one subject (20/40). Audio thresholds were determined for 0.25–8 kHz and exceeded the 20 dB threshold only in one subject (30 dB for 4–8 kHz in left ear). All participants were right handed (Edinburgh Handedness inventory), had at least average intelligence (full scale >100 on the Wechsler Abbreviated Scale of Intelligence) and had no history or presence of neurological or psychiatric illness. The study was approved by the Institutional Review Board at the New York State Psychiatric Institute.

Stimuli

Visual stimuli were color photographs of objects presented on a black background and auditory stimuli were sound clips of the same objects, both obtained from the Multimodal Stimulus Set [Schneider et al., 2008] from which we selected 72 objects that in a pilot experiment had high ratings for auditory-visual correspondence. The set included 36 animal or vehicle objects, and 36 objects belonging to other categories (musical instruments, sport objects, tools, weapons, and everyday objects). Stimuli were presented visual-only, auditory-only, or audiovisually (always congruent and simultaneous). Visual stimuli were presented using a stereoscopic goggle system (model VOU-RTC2k, 2004, Resonant Technology Inc, Northridge, CA) that provided a total (left to right) visual field of 30°. Auditory stimuli were presented using MR-compatible stereo headphones (Resonant Technology Inc, Northridge, CA) that attenuate scanner noise up to 30 dB. Participants did not wear earplugs as this could have induced unwanted left-right differences in sound intensity.

Pictures

Pictures were presented individually at foveal or peripheral (left) locations on the screen. Because more cortical tissue is devoted to foveal than to peripheral portions of the visual field (cortical magnification), and because participants had to be able to detect and recognize both the central and peripheral pictures equally well, we accounted for cortical magnification and its effect on performance [see also Meinenbrock et al., 2007]. The details of this correction are given in the Supporting Information Materials and

Methods. The optimized parameters in our set-up were: (1) foveal pictures with a visual angle of 1.27°, and (2) peripheral pictures at center eccentricity of 12.4° and a visual angle of 5.3°.

Sounds

Sounds were all stereo 16-bits wave files with a sampling rate of 22 kHz, matched for intensity (root mean square power) using Praat software (Boersma & Weenink, University of Amsterdam, The Netherlands). Apparent sound location was manipulated by inserting 1 ms of silence at the beginning of the acoustic signal in the right-sided channel. This creates an inter-aural time difference (ITD) that is perceived as a completely left-lateralized sound [von Kriegstein et al., 2008], which we verified in a pilot experiment.

Task Design

Stimuli were presented using a $2 \times 2 \times 3$ design (Fig. 1A) with factors Task (speed, accuracy), stimulus Eccentricity (central, peripheral), and sensory Modality (visual, auditory, audiovisual). The factors Task and Eccentricity were blocked and Modality was randomized within blocks. The Task factor was blocked (and alternated every block) because performing two different tasks in rapid alternation would be exceedingly difficult. The Eccentricity factor was blocked (and alternated every two blocks) to facilitate the ability of participants to fixate on the center of the screen throughout the entire experiment. In the accuracy task, participants were instructed to decide (by a two-choice button press) as accurately as possible whether each successive picture or sound (or combination) belonged to the “animal or vehicle” categories or to another category (“other”). Similar to tasks used in other studies on audiovisual object integration [Beauchamp et al., 2004a; Werner and Noppeney, 2010], this task forced participants to process the details of the stimuli up to the category level. In the speed task, participants were instructed to press a button as fast as possible when they detected (by seeing and/or hearing) an object, regardless of its identity. Participant responses were recorded using an MR-safe custom-made button box.

E-prime 2.0 (Psychology Software Tools, Inc, Sharpsburg, PA) was used for stimulus presentation. The timing of events is schematically depicted in Figure 1B. The first fixation period (20 s) was triggered by the first radiofrequency (RF) pulse from the scanner. Subsequently, 12 task blocks were presented, with each block followed by a rest period of 5 s, during which participants could close or move their eyes, and a fixation (baseline) period of 12 s that was signaled with a 500 ms audiovisual cue (a low-pitched beep and the word “focus!”). The task was signaled by a second cue (a high-pitched beep with either the words “what is it?” or “be fast!”). Stimuli were presented for 400 ms, with stimulus onsets synchronized to scanner

Figure 1.

Experimental design and timing of events. **A**) The fully balanced $2 \times 2 \times 3$ design of the fMRI experiment with factors task (speed, accuracy), stimulus (peripheral, central) and modality (auditory/A, visual/V, audiovisual/AV). **B**) Timing of events in the mixed fMRI design. Task and Stimulus Eccentricity were blocked, Modality was randomized within blocks. Twelve stimuli were

presented per block (four of each modality) using a jittered stimulus onset asynchrony (SOA) of 2.5 to 6.25 s (=2–5 scanning repetition times/TRs). Between blocks, after a rest period of 5 s (during which participants did not have to fixate), the fixation baseline period was signaled (AV Cue 1). Task instruction was cued 2.5 sec before the next task block (AV Cue 2).

pulses. Stimulus onset asynchrony (SOA; between subsequent objects, not between V and A of the same object) was jittered between 2.5, 3.75, 5, and 6.25 s (=2–5 scanning repetition times/TRs).

Twelve objects were presented in each block. The modalities (four of each), SOAs (three of each), and categories (six “animal- or vehicle” and six “other”) were randomized. Block order was counterbalanced across runs.

In total, four runs (of 12 blocks each) were obtained for each participant. Participants performed a short practice run before entering the scanner, to ensure that both task instructions and cues were understood, and to train them to fixate their eyes on the center of the screen (as they were instructed to do at all times).

Eye-Tracking

Eye position was monitored during scanning as saccades to the peripheral pictures would confound the eccentricity manipulation. Eye movements were monitored using an eye-tracker device (model RTC2K, 2004, Resonant Technology Inc, Northridge, CA) custom-mounted on a glasses frame to create stable positioning of the camera. Eye position was recorded using ViewPoint EyeTracker software (version 2.7, Arlington Research, Inc, Scottsdale, AZ) at a frequency of 30Hz. Prior to starting the experiment, the eye-tracker was calibrated using a standard procedure provided by the ViewPoint software.

Scanning Protocol

Four functional runs (768 volumes each) and an intra-session anatomical image were acquired for each participant on a 3T scanner (GE Signa, Wisconsin) with a quadrature head coil (GE, Wisconsin). The task was then practiced once more during a short fMRI run to test experimental sound volume. The functional runs were acquired using the following pulse sequence: TR 1,250 ms, TE 27 ms, Flip angle 68°, matrix 64×64 , FOV 224×224 mm, slice thickness 4.0 mm, gap 1 mm. Twenty-two axial slices were scanned, positioned in such a way that occipital and temporal lobes fitted optimally. The within-session anatomical scan acquired in each subject was a Fast Spoiled Gradient-Recalled-Echo (sagittal, $1 \times 1 \times 1$ mm isotropic voxels, TE minimal, matrix 256×256).

Data Analysis

Eye-tracking data

See Supporting Information Materials and Methods for analysis procedures. Removing first-picture trials from each block eliminated the majority of saccades (Supporting Information Fig. S1a). We therefore removed all first-picture trials from the subsequent behavioral and fMRI data analyses.

Behavioral data

As an index of multisensory integration, we calculated audiovisual (AV) advantage scores as the difference between the AV reaction times (RTs) and the fastest of the unisensory RTs (A or V), and between the AV hit rates (HR) and the highest unisensory HR (A or V). Only correct responses were included in calculating RTs. Statistical

analyses were performed using repeated measures ANOVAs using AV-advantage scores as dependent variable with factors Task and Eccentricity as independent variables. To test whether RT AV-advantage effects exceeded the facilitation that would be predicted when the two inputs would be processed independently, we calculated Race Model Inequality (RMI) scores [Miller, 1982]. This analysis distinguishes whether the AV effects can be explained by race models (i.e., independent processing) or by coactivation, the latter indicating neural multisensory integration. For each subject and condition, RT's within that subject's valid range (100 ms – (mean RT + 2 standard deviations)) were binned into 21 quantiles from 0 to 100% in 5% increments (1,5,10,...,95,100%). Next, the cumulative probability for each condition in each bin was calculated and averaged across subjects to get the average cumulative probability (CP) distributions. We then calculated the race model prediction using the Miller inequality equation ($[CP_A + CP_V] - [CP_A \times CP_V]$) at each bin. The values predicted by the race model were then subtracted from the audiovisual CP values at each bin to get the Race Model Inequality (RMI) scores, which were tested against zero using one-sample *t*-tests. *P* values are given at uncorrected and Bonferroni-corrected thresholds ($P < 0.05$ and $P < 0.0025$, respectively). Positive RMI scores indicate deviation from the race model prediction and therefore suggest neural integration of the inputs. RMI scores for the different eccentricity conditions were tested against each other using paired-samples *t*-tests. For a similar approach, see [Laurienti et al., 2004; Molholm et al., 2002; Murray et al., 2001; Romei et al., 2007]. Although the RMI analysis is mainly relevant for the speeded detection task as the subjects were instructed to respond as fast as possible in this task only, we also performed the same analysis for the accuracy task for completeness. Finally, we performed repeated measures ANOVAs using RT and hit rate and the factors Task, Eccentricity and Modality. Further analyses of the behavioral data that take individual differences in visual (central vs. peripheral) performance into account are explained in the Supporting Information Materials and Methods.

fMRI time-series

BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands) [Goebel et al., 2006] was used for fMRI data analysis. Preprocessing included slice time correction, linear trend removal, temporal high pass filtering (2 sines/cosines per run) and three-dimensional motion correction (sinc interpolation). Functional slices were co-registered to the anatomical volume using position parameters from the scanner and intensity-driven fine-tuning, and transformed into Talairach space, interpolating to a resolution of $3 \times 3 \times 3$ mm³. The volume time-course data were spatially smoothed using a Gaussian kernel of 5 mm Full Width at Half Maximum. For data presentation, an averaged anatomical volume was created from the 18 individual anatomical volumes.

In our search for contextual effects on multisensory integration (MSI) circuitry, we performed a multistep analytical approach. In a first step, we tested for MSI on a whole-brain level using multi-subjects random-effects general linear models (GLM) combined with the “max criterion.” According to this criterion, the multisensory activation should be stronger than the most effective unisensory condition. Although the inference is not unequivocal with regard to neural integration (see Discussion), it has been suggested as the optimal criterion to infer multisensory integration from fMRI data [Beauchamp, 2005; Doehrmann and Naumer, 2008]), especially for whole-brain analyses [Goebel and van Atteveldt, 2009]. However, it should be noted that there is still some debate about which criterion to use [Stevenson et al., 2009; Werner and Noppeney, 2010], see also Discussion. In a second step, we performed regions-of-interest (ROI)-based analyses to investigate Task effects on MSI specific to the central vs. peripheral representations, as the central/peripheral distinction was core to our predictions. A similar approach was taken as Meinenbrock et al. [2007] used for a position-resolved analysis of audiovisual integration effects. Using single-subject GLM analyses we contrasted all peripheral pictures/sounds (peripheral V/A + AV, collapsed over Task) vs. all central pictures/sounds (central V/A + AV, collapsed over Task). Using the sound trials, this individual analysis did not reveal any eccentricity-specific ROIs in auditory cortex. Using the picture trials, this analysis resulted in individual retinotopic ROIs in 16/18 subjects.

Whole-Brain Analyses

For the first level analysis, all experimental conditions were modeled as separate predictors in each subject. Predictor time-courses were adjusted for the hemodynamic response delay by convolution with a double-gamma hemodynamic response function. The resulting GLM contained 12 predictors of interest per subject: task (2) × eccentricity (2) × modality (3). Additional predictors were created for cue 1 and cue 2, rest period, and the first-picture trials of each block were modeled separately as a correction for eye movements. The six motion-correction time-courses were included as confound predictors. We tested for MSI effects by second-level random-effects conjunction of the Modality contrasts (AV vs. A) ∩ (AV vs. V), this corresponds to the aforementioned “max criterion.” We performed this test collapsed over task × eccentricity factors as well as within each of the four task × eccentricity combinations separately. To correct for multiple comparison, maps thresholded at an initial voxel-level *P* value were submitted to a whole-data correction criterion based on the estimate of the map’s spatial smoothness and on iterative Monte Carlo simulations for estimating cluster-level false-positive rates [Forman et al., 1995; Goebel et al., 2006]. After 1,000 iterations, the minimum cluster-size corresponding to a corrected false positive probability of <0.05 was applied to the statistical maps.

From the resulting maps, we extracted individual beta-estimates (BE) for all 12 conditions and used these to calculate individual MSI values for the four task × eccentricity combinations by subtracting the highest unisensory BE from the multisensory BE ($MSI = AV - \max[A, V]$). These MSI values were entered in an ANOVA with factors task and eccentricity.

ROI-Based Analyses

We performed ROI-based analyses to investigate MSI specific to the central versus peripheral representations. Moreover, as the whole-brain analysis was performed on a group level, we wanted to complement this approach with an analysis using individually defined ROIs, to be more sensitive to effects taking place in functional ROI’s that might not overlap after spatial normalization. For the sounds, the contrast (central A + AV) versus (peripheral A + AV) did not reveal any eccentricity-specific ROIs in auditory cortex. In the visual cortex, we defined a right hemisphere “Peripheral” and “Central,” and a left hemisphere “Central” ROI (as the central pictures comprised the foveal visual field on both sides) in individual participants using the contrast of all peripheral pictures (peripheral V + AV, collapsed over task) versus all central pictures (central V + AV, collapsed over task). From these individual retinotopic ROIs, we extracted the BEs for all 12 conditions and used these to calculate MSI values for the four task × eccentricity combinations ($AV - \max[A, V]$).

Analysis of Subgroups and BOLD Dynamics

We performed both whole-brain and ROI-based analyses also while considering individual variation in visual (central vs. peripheral) performance, as we did for the behavioral analyses, these methods are explained in the Supporting Information Materials and Methods. To examine the dynamics of the fMRI response, we reconstructed the BOLD response from the time-courses in the clusters revealed by the whole-brain and ROI-based analyses using deconvolution analyses. The BOLD response for each condition was estimated using a GLM with a set of 16 shifted stick predictors for each event, which correspond to estimating the BOLD response for 20 s ($16 \times TR$) post-event. We quantified differences in BOLD response dynamics using paired *t*-tests on the beta-estimates for AV versus V (in visual cortex) and AV versus A (in auditory cortex) during the peak of the BOLD response (individually defined) and during the decay phase (average beta-estimate over time points 5–7).

RESULTS

Behavioral and Eye Tracking Results

Monitoring of eye position disclosed low saccade counts (Supporting Information Fig. S1a), and revealed a

significant main effect of modality ($F_{(1,10)} = 5.1$; $P < 0.05$), deriving from fewer saccades to audiovisual (AV) trials compared with visual (V) trials. As most of the saccades occurred to the first picture-trial (V or AV) of a block we removed first picture-trials (right graph). Based on these findings, we removed all first picture-trials in all 18 subjects from the subsequent behavioral and fMRI data analyses.

The effects of Task and Eccentricity on multisensory integration (MSI) were tested using reaction time (RT) and Hit Rate (HR) AV-advantage scores (Fig. 2A). HR AV-advantage scores (accuracy task only) were significantly positive for both eccentricities (Peripheral: $t_{17} = 2.2$, $P < 0.05$; Central: $t_{17} = 2.2$, $P < 0.05$). RT AV-advantage scores revealed a main effect of Eccentricity ($F_{(1,17)} = 4.2$, $P < 0.05$) and a trend towards an interaction of Task and Eccentricity ($F_{(1,17)} = 3.6$, $P < 0.07$). Post hoc comparisons revealed significant higher AV-advantage for central versus peripheral stimuli in the accuracy task ($t_{17} = 2.2$, $P < 0.05$). Testing all AV-advantage scores against zero (i.e., no advantage) confirmed a significant speed advantage for bimodal stimulus presentation in all conditions except for peripheral stimuli in the accuracy task (Peripheral Accuracy: $t_{17} = 1.00$, $P = 0.34$; Peripheral Speed: $t_{17} = 4.0$, $P < 0.001$; Central Accuracy: $t_{17} = 3.4$, $P < 0.003$; Central Speed: $t_{17} = 6.2$, $P < 0.0001$).

Figure 2B shows the Race Model Inequality results (the corresponding cumulative probability distributions including the race model prediction curves are shown in Supporting Information Fig. S1b). This analysis was mainly relevant for the speed task but the results for the accuracy task are shown as well. For the speed task, the race model prediction was violated for the fastest quarter of the responses (<25%), and also for some later quantiles (between 50 and 65%) for the peripheral stimuli. In these bins, the AV responses were faster than predicted by simple probability summation of A and V. Moreover, for the fastest 5 to 10%, RMI was larger for the peripheral than for the central stimuli. For the accuracy task, the race model was not violated by the AV conditions. However, the RMI was significantly larger for central than for peripheral stimuli at the 10% RT bin.

Figure 2C shows RTs and HRs during scanning. The RTs for the accuracy task showed a significant main effect of Modality ($F_{(2,34)} = 51.3$; $P < 0.001$) and an interaction between Eccentricity and Modality ($F_{(2,34)} = 5.6$; $P < 0.01$). Post hoc comparisons (paired t -tests) between AV and unisensory RT's revealed significant differences in all pairs except V versus AV for the peripheral stimuli (Peripheral: A vs. AV $t_{17} = 11.1$, $P < 0.001$; V vs. AV $t_{17} = 1.0$, $P = 0.33$. Central: A vs. AV $t_{17} = 5.0$, $P < 0.001$; V vs. AV $t_{17} = 4.2$, $P < 0.001$). HRs for the accuracy task showed main effects of Eccentricity ($F_{(1,17)} = 10.3$, $P < 0.005$) and Modality ($F_{(2,34)} = 69.5$, $P < 0.001$) and post hoc comparisons revealed significant differences between AV and both unisensory HR's (Peripheral: AV vs. A $t_{17} = 11.3$, $P < 0.001$; AV vs. V $t_{17} = 2.2$, $P < 0.04$. Central: AV vs. A $t_{17} = 9.7$, $P < 0.001$; AV

vs. V $t_{17} = 2.3$, $P < 0.03$). RTs for the speed task showed a main effect of Modality ($F_{(2,34)} = 80.0$; $P < 0.001$) and of Eccentricity ($F_{(1,17)} = 4.8$; $P < 0.05$), and post hoc comparisons revealed significant differences between AV and both unisensory RT's (Peripheral: A vs. AV $t_{17} = 12.29$, $P < 0.001$; A vs. V $t_{17} = 6.91$, $P < 0.001$; V vs. AV $t_{17} = 2.99$, $P = 0.008$. Central: A vs. AV $t_{17} = 15.2$, $P < 0.001$; A vs. V $t_{17} = 4.0$, $P < 0.001$; V vs. AV $t_{17} = 5.1$, $P < 0.001$). For comparison, RT data for the same tasks outside the scanner are shown in Supporting Information Figure S1c.

The revealed effects of Eccentricity were unexpected, as the enlarged peripheral pictures (to correct for cortical magnification) should have matched the performance for peripheral and central pictures. Instead, the effects suggest that peripheral trials were easier. Supporting Information Figure S1d shows the results of the subgroups analysis of participants based on central vs. peripheral visual RTs. The RTs by subgroup show that the eccentricity matching was only suboptimal in a subset of participants (the "non-matched" subgroup) who showed worse performance for the central (V and AV) trials. The other subset ("matched" subgroup) showed the intended matched performance for peripheral and central picture-trials (V and AV). Saccade counts by subgroup did not show an effect of group ($F_{(1,9)} = 0.6$, $P = 0.46$).

fMRI Results—Whole-Brain Analyses

Figure 3A shows the results of the whole-brain, random-effects, general linear model (GLM) analysis with predictors for sensory Modalities (A, V, and AV) collapsed over Task and Eccentricity. We defined MSI by the conjunction of (AV vs. A) $\cap$ (AV vs. V) (see Materials and Methods). The map in Figure 3A is corrected for multiple comparisons at the cluster-level ($P < 0.05$; voxel-level $t_{17} = 4.6$, $P < 0.0003$, cluster-size threshold eight voxels) and reveals activation exclusively in auditory cortex bilaterally, in Heschl's gyrus and Heschl's sulcus. The location seems to correspond to low-level auditory cortex [Hackett et al., 2001], but note that this is a group result projected onto the group average anatomical image. The bar graphs show the fMRI response (beta estimates) combined for the left and right auditory activation clusters, as these did not reveal any differences (no main effect of hemisphere: $F_{(1,17)} = 2.9$; $P < 0.1$). The bar graphs in Figure 3A indicate that MSI differs across the Task and Eccentricity conditions, and that unisensory activation is task-dependent (speed task produced greater visual and reduced auditory activation). To test the first of these observations, we entered MSI values (AV – max[A,V]) in an ANOVA with factors Task and Eccentricity, revealing a significant interaction of Task and Eccentricity ($F_{(1,17)} = 4.89$, $P < 0.041$), no main effects. Post hoc paired t -tests revealed that MSI differed for peripheral and central stimuli during the accuracy task ($t_{17} = 3.15$, $P < 0.005$), but not during the speed task ($t_{17} = 0.66$, $P = 0.52$). The other comparisons did not

Figure 2.

Behavioral results. **A**) Above: reaction time (RT) advantage scores for audiovisual (AV) trials (subtraction of the AV RT from the minimal RT for A or V). Below: hit rate (HR) advantage scores for AV trials (subtraction of maximal A or V HR from the AV HR). Advantage scores are plotted separately for the speed task (dotted line) and accuracy task (solid line). **B**) Results from the Race Model Inequality (RMI) analysis. RMI curves (AV probability distribution minus race model prediction) for the speed task (above) and for the accuracy task (below). Light grey lines: peripheral stimuli; dark grey lines: central stimuli.

Significance levels indicate uncorrected P values (<0.05) and Bonferroni-corrected P values (<0.0025). Stars (*) give significance levels for RMI versus zero t -tests, section signs (§) for RMI-peripheral versus RMI-central t -tests. See Supporting Information Figure S1b for the corresponding cumulative probability distributions for each condition. **C**) RT and hit rates plotted for the accuracy task and speed task. Dark grey bars (A): auditory trials; medium grey bars (AV): audiovisual trials; light grey bars (V): visual trials. Error bars indicate standard error of the mean (SEM). See also Supporting Information Figure S1.

reveal significant differences (central-accuracy vs. speed: $t_{17} = 1.90$, $P = 0.075$; Peripheral-Accuracy vs. Speed: $t_{17} = 0.69$, $P = 0.50$). Task-dependent unisensory activation

was tested using the beta estimates for unisensory activation in an ANOVA with factors Modality (visual, auditory), Task, and Eccentricity. As we detected a significant

A

B

Figure 3.

effect of Modality ($F_{(1,17)} = 26.59$, $P < 0.00$), we tested the factors Task and Eccentricity separately for visual and auditory activations. We found a main effect of Task for both (Visual: $F_{(1,17)} = 7.27$, $P < 0.015$; Auditory: $F_{(1,17)} = 7.74$, $P < 0.013$), no main effect of Eccentricity or Task $\times$ Eccentricity interaction.

The MSI analysis collapsed over Task and Eccentricity could be biased toward the Task and Eccentricity levels that produce the strongest MSI activation in those voxels, and opposite effects in different Task/Eccentricity conditions could have cancelled each other. Therefore, we also performed the conjunction of (AV vs. A) $\cap$ (AV vs. V) separately for each Task/Eccentricity combination (Fig. 3B). The maps in Figure 3B are corrected for multiple comparisons at the cluster-level ($P < 0.05$; voxel-level $t_{17} = 3$, $P < 0.008$, cluster-size thresholds per map: 25 (Accuracy-Peripheral), 16 (Accuracy-Central), 18 (Speed-Peripheral), 19 (Speed-Central)). In addition to the effects in the auditory cortex, a cluster in the posterior ascending segment of the STS (green map; Talairach coordinates $-48x$, $-58y$, $31z$) showed stronger MSI for the accuracy task (main effect of Task ($F_{(1,17)} = 6.24$, $P < 0.023$), no main effect of Eccentricity or Task $\times$ Eccentricity interaction).

fMRI Results—ROI-Based Analyses

The individual ROI-based analyses aimed at directly comparing MSI in foveal/central versus peripheral representations. Also, it was used to complement the whole-brain analysis as it is less vulnerable to spatial averaging confounds. We did not find eccentricity-specific ROIs in the auditory cortex, so the planned ROI-based analysis was not possible in auditory cortex. In the visual cortex, 2 of the 18 participants did not show retinotopic ROIs and so were excluded from the ROI-based analysis. The central ROIs in the right occipital cortex were more robust and consistent across participants compared with those in the left hemisphere, so we restricted further analysis to peripheral and central ROIs in the right hemisphere (Fig. 4). The ROIs were selected at a similar cluster size (at individual thresholds) of approximately 400 voxels (anatomical resolution; mean size = 401, SEM = 3,

corresponding functional resolution ≈ 15 voxels). The mean Talairach coordinates were ($27x$, $-87y$, $9z$) for central, and ($18x$, $-67y$, $-13z$) for peripheral ROIs. The bar graphs in Figure 4 show that most MSI effects were in the direction opposite those in the auditory cortex: weaker fMRI activation for AV compared with V stimulation. The ANOVA of MSI values (AV – max[A,V]; indeed mostly negative) revealed a main effect of Task ($F_{(1,15)} = 12.42$, $P < 0.003$) and a trend toward an interaction of Task $\times$ Eccentricity ($F_{(1,15)} = 3.82$, $P < 0.07$), no main effect of Eccentricity.

fMRI Results—Comparison of Effects in Auditory and Visual Cortex

To compare the MSI effects in auditory and visual cortices side-by-side, we plotted the MSI values from the revealed auditory (see Fig. 3A) and visual (see Fig. 4) clusters in the same graph (Fig. 5A). We note that the clusters from which the MSI values were computed were defined differently: the auditory clusters were defined by a whole-brain group-based GLM, whereas the visual clusters were defined individually and using a different statistical test. Keeping this in mind, the profiles of Task and Eccentricity effects on MSI in both sensory cortices show interesting similarities and differences. Analysis of variance of the MSI values with factors Cortex (visual, auditory), Task, and Eccentricity revealed a significant three-way interaction ($F_{(1,15)} = 6.82$, $P < 0.020$). To understand this interaction better, we compared all means with each other using paired t -tests. For the accuracy task, MSI for peripheral versus central stimuli differed within sensory cortex, but was not different across cortex (peripheral-visual-cortex vs. peripheral-auditory-cortex: $t_{15} = 1.30$, $P = 0.21$; central-visual-cortex vs. central-auditory-cortex: $t_{15} = 1.07$, $P = 0.30$). For the speed task, MSI for peripheral versus central stimuli did not differ within sensory cortex, but did differ across cortex (peripheral-visual-cortex vs. central-visual-cortex: $t_{15} = 0.56$, $P = 0.58$; peripheral-auditory-cortex vs. central-auditory-cortex, $t_{17} = 0.60$, $P = 0.56$; peripheral-visual-cortex vs. peripheral-auditory-cortex: $t_{15} = 5.35$, $P < 0.00$; central-visual-cortex vs. central-

Figure 3.

Whole-brain random-effects GLM analysis of multisensory integration. **A**) Results of the conjunction of (AV vs. V) $\cap$ (AV vs. A), collapsed over Task and Eccentricity. The resulting map (left panel) is corrected for multiple comparisons at the cluster-level ($P < 0.05$). Middle panel: extracted beta estimates from the superior temporal clusters shown in the left panel, averaged across participants and plotted for each condition separately (combined for left and right hemisphere clusters). A = auditory, AV = audiovisual, V = visual. Right panel: upper graph displays the multisensory interaction (MSI) values (AVbeta –

Max[A,V]beta) for the two tasks as a function of eccentricity; bottom graph displays the unisensory (auditory/A and visual/V) activation in the same superior temporal clusters for the two tasks as a function of eccentricity. **B**) Results of the conjunction of (AV vs. V) $\cap$ (AV vs. A), separate for each Task $\times$ Eccentricity combination. The resulting four maps (see color coding in figure) are corrected for multiple comparisons at the cluster-level ($P < 0.05$). MSI values (AVbeta – max[A,V]beta) from the green map (accuracy task/peripheral stimuli) were plotted for each task as a function of eccentricity (right panel).

Figure 4.

Peripheral and central individual ROIs in the visual cortex. fMRI responses in individually defined retinotopic areas in the visual cortex. Left panel shows the peripheral (blue-green) and central (orange-yellow) ROIs in six representative participants. Please note that these slices are selected to optimally show the retinotopic ROIs in the right hemisphere, the ROIs on the left might

have been more representative in lower/higher slices. Middle panel: beta estimates extracted from the individual ROIs plotted for each condition (A = auditory; AV = audiovisual; V = visual). Right panel: multisensory interaction values ($AV\beta - \max[A, V]\beta$) for the two tasks as a function of eccentricity.

auditory-cortex: $t_{15} = 4.31$, $P < 0.001$. As already noted, MSI in visual and auditory cortex was opposite in direction, with enhancement in the auditory cortex and suppression in the visual cortex. Blood Oxygenation Level Dependent (BOLD) response dynamics in the same clusters (Fig. 5B) show a significantly higher peak for AV versus A ($P < 0.05$) but no differences during the BOLD decay phase ($P < 0.18$) in the auditory cortex. In contrast, in the visual cortex, the AV BOLD response never peaks as high as the visual one ($P < 0.03$), and is clearly lower during the decay phase ($P < 0.006$), indicating that it returns to baseline faster in the AV condition.

fMRI Results—Individual Differences in Task and Eccentricity Effects

As their classification was based on (peripheral vs. central) visual processing, the subgroups presented in Supporting Information Figure S1d may have produced different MSI profiles in the visual cortex. A whole-brain analysis with Subgroup as between-subjects factor (see

Supporting Information Materials and Methods) revealed a Subgroup $\times$ Modality interaction in only one area, a higher-level (i.e., non-retinotopic) visual cluster located in the right hemisphere fusiform gyrus, lateral/anterior to the retinotopic ROIs (Supporting Information Fig. S2). This cluster showed (negative) MSI only for the speed task (main effect of Task, $F_{(1,16)} = 10.12$, $P < 0.01$); however, the MSI effect appears to be absent for central stimuli in the “non-matched” subgroup. To directly examine subgroups differences in low-level visual cortex, MSI values derived from the individually defined retinotopic ROIs (see Fig. 4) were plotted separately for the two subgroups (Supporting Information Fig. S2, lower panel). These split plots show different MSI profiles for the two subgroups (Task $\times$ Subgroup interaction: $F_{(1,14)} = 5.81$, $P < 0.03$; derived from an effect of Subgroup for the accuracy task $F_{(1,14)} = 6.80$, $P < 0.02$ that was absent for the speed task). The subgroups again appear to differ especially for the central stimuli during the accuracy task; however, the Subgroup $\times$ Eccentricity interaction did not reach significance.

Figure 5.

Comparison of Task and Eccentricity effects on multisensory integration and BOLD dynamics in auditory vs. visual cortex. **A**) The multisensory integration (MSI) profiles from the auditory clusters and visual retinotopic ROIs plotted in the same graph to compare Task and Eccentricity effects on MSI. Dark-gray lines: MSI values from auditory cortex; light-gray lines: MSI values from visual cortex. Dotted lines: speed task; solid lines: accuracy task. **B**) Deconvolution plots for the speed task in the auditory and visual clusters. Dark-gray lines: estimated BOLD response to auditory trials; medium-gray lines: to audiovisual trials; light-gray lines: to visual trials. TR = scanning Repetition Time (= 1.25 s).

DISCUSSION

In this study, we addressed the hypothesis of “functional specialization” of different circuits for low-

level multisensory integration. In particular, we tested the predictions that neuronal architectures for multisensory integration (MSI) are specialized for speed in peripheral detection (low-level sensory cortex, peripheral fields) and for accuracy in foveal/central discrimination (higher-order association cortex and low-level central sensory fields). While our findings support some of these predictions, they also point to a more flexible use of neural circuitry for MSI than the hypothesized speed/peripheral versus accuracy/central specialization. When accuracy was required, both low-level auditory and visual cortices expressed multisensory enhancement for peripheral stimuli but no enhancement, or even a weak suppression, for central stimuli. When speed was required, MSI was robust and independent of stimulus eccentricity, but its effects were in opposing directions in auditory (enhancement) and visual (suppression) cortices. Moreover, individual differences in central vs. peripheral visual behavioral performance seemed to influence MSI profiles further in visual cortex. Finally, a higher-level association region in posterior STS/angular gyrus preferentially integrated audiovisual information during the accuracy task. Before discussing the revealed fMRI effects of Task and Eccentricity on MSI, we will first consider the implications of the eye-tracking and behavioral results.

Multisensory Effects on Task Performance

The low average saccade counts (Supporting Information Fig. S1a) provide assurance that during the majority of peripheral trials, participants were fixating on the center of the screen and that the peripheral pictures thus fell on the intended peripheral location on the retina. Importantly, the behavioral results during scanning (Fig. 2) show significant performance advantages for audiovisual stimulus presentation relative to unimodal presentation. Audiovisual hit rates (HR; accuracy task) were increased for all Task $\times$ Eccentricity conditions, and reaction times (RT) were reduced for both eccentricities in both tasks, except for the peripheral stimuli during the accuracy task, which is conform our prediction that accuracy demands enhance integration specifically for central inputs.

The Race Model Inequality (RMI) test for the speed task supports the interpretation that the AV-advantage effects are due to multisensory integration as the AV responses are faster than would be predicted by simple probability summation. Interestingly, the AV responses deviated from the race model more strongly for the peripheral stimuli compared to the central stimuli. This effect is in the predicted direction: speed demands were predicted to enhance integration mainly for peripheral inputs. Note that this eccentricity effect was not found for the mean RTs, which indicates the importance of examining the probability summations for bimodal inputs in addition to mean RTs. The AV responses did not violate the race model for the accuracy task, which is not unexpected

given that the task instruction focused the participants on being accurate rather than being fast. Therefore, the HR- and RT-advantage effects in this task are strong indicators multisensory integration, especially the increased HR for AV inputs.

There are two additional notable observations from the behavioral results. One is that the cortical magnification correction seems to have effectively matched peripheral and central visual performance only in a subset of the subjects (Supporting Information Fig. S1d); this will be discussed in the section Individual Variation in Multisensory Integration. The other is the performance advantage provided by visual-alone compared to auditory-alone stimuli, which suggests that the auditory input is less robust and that vision is the dominant modality for either task in the current experimental set-up. This might seem at odds with cortical response latencies [Musacchia and Schroeder, 2009] and faster simple reaction times for sounds versus lights [Barnett-Cowan and Harris, 2009; Welford, 1980; Woodworth and Schlosberg, 1954], however, the sounds used in this study are complex environmental sounds, for which the detectable/identifying acoustic information accumulates over time, in contrast to simple tones with fast onsets. The normative study of the stimulus set used here also found considerably lower hit rates and longer RTs for categorization of the sounds compared with pictures [Schneider et al., 2008]. Other studies using complex visual and auditory object stimuli found similar RT length and differences between sounds and pictures [Beauchamp et al., 2004a]. The acoustic scanner noise may have additionally influenced task difficulty for the sounds. However, behavioral data outside the scanner (Supporting Information Fig. S1c), as well as the results in Schneider et al. [2008], show auditory versus visual RT differences in the same direction, but in the absence of scanner noise.

Task and Stimulus Effects on fMRI Integration Indices

As mentioned above, the fMRI results partly support the predicted speed/peripheral versus accuracy/central functional specialization within low-level and higher-level cortices. Below, we will first discuss the task and eccentricity effects on the used MSI indices in low-level auditory and visual cortices and higher-level STS, and will focus on the potential meaning of the deviations from our predicted neural architecture next.

In auditory cortex, MSI strength showed an interaction of task and stimulus eccentricity (Fig. 3A). MSI during the speed task was independent of stimulus eccentricity, whereas only the peripheral stimuli produced MSI during the accuracy task. Furthermore, activation of auditory cortex by unisensory visual stimuli was stronger during the speed task. Both findings are consistent with the predicted functional specialization of low-level sensory cortex in integrating peripheral inputs under speeded

circumstances, but also indicate a more flexible recruitment of low-level auditory cortex. Two notable deviations from the predicted specialization are (1) cancellation of eccentricity effects during the speed task; and (2) a peripheral bias in MSI during the accuracy task, which was actually predicted for the speed task.

In visual cortex, we observed similar effects: a main effect of task and a trend towards interaction of task and eccentricity on MSI (Fig. 4). The task effect is expressed as stronger (negative) MSI for the speed task. These findings also partly correspond to the predictions for low-level cortex (i.e., stronger integration during speed demands), but again indicate a more flexible use of the circuitry than the predicted foveal/accuracy vs. peripheral/speed specialization. Moreover, as most of the MSI effects were suppressive in visual cortex, integration here may differ qualitatively from integration in auditory cortex. Although the suppressive MSI in visual cortex could indicate interference by the less robust auditory input, audiovisual vs. visual presentation nevertheless improves behavioral performance. Therefore, MSI in the visual cortex likely reflects a beneficial combining of AV inputs. The deviations from our predicted specialization are (1) cancellation of eccentricity effects during the speed task, as was the case in auditory cortex; (2) suppressive integration effects for the speed task; and (3) although the Task $\times$ Eccentricity interaction did not reach significance, the MSI profile suggests the same peripheral bias in MSI during the accuracy task as in auditory cortex.

The whole-brain survey for multisensory integration revealed one cluster in higher-order cortex, located on the posterior ascending segment of the superior temporal sulcus (STS) [Ochiai et al., 2004] and possibly angular gyrus. This cluster preferentially integrated audiovisual stimuli during the accuracy task (main effect of task, Fig. 3B), as predicted for the functional specialization of feedback integration pathways. However, the predicted stronger integration of central information, relative to peripheral, was not confirmed. Although this region is more posterior (and superior) than typically reported posterior STS integration effects [Amedi et al., 2005; Hein and Knight, 2008; Hocking and Price, 2008], MSI in more posterior locations on STS has been found by other studies as well [Kreifelts et al., 2009; Taylor et al., 2006], including one that presented audiovisual objects at similar eccentricities [Meienbrock et al., 2007].

Contextual Control of MSI: Deviations From the Predicted Functional Specialization

Side-by-side comparison of multisensory integration indices in auditory and visual cortex (Fig. 5) revealed a comparable task-dependent integration pattern in these two cortices, although integration was expressed in opposite direction in the speed task. Multisensory interactions have been shown to be strongly task-dependent before

[Fort et al., 2002; Giard and Peronnet, 1999; Van Atteveldt et al., 2007b; Wang et al., 2008]. The different pattern of behavioral results for the two tasks (Fig. 2) further supports task-dependent integration processes: audiovisual input in the categorization task improves detailed processing for recognition purposes (hit rates) more than it speeds up processing, whereas it robustly speeds up detection in the speed task as confirmed in the race model analysis.

During the accuracy task, both low-level cortices expressed the same effect of eccentricity, namely multisensory enhancement for peripheral stimuli and no enhancement or a weak suppression for central stimuli. Stronger MSI for peripheral vs. central stimuli is consistent with eccentricity-dependent anatomical projections [Falchier et al., 2002, 2010], but was predicted specifically for speed tasks; the accuracy task was predicted to produce MSI more strongly for the central stimuli. Stronger integration for peripheral stimuli thus seems more general than only for speeded detection. Other studies also report stronger peripheral MSI for tasks other than speeded detection, such as size judgment [Takeshima and Gyoba, 2013].

When speed was the main goal, eccentricity effects seemed to be overruled in both auditory and visual cortex, as the same MSI strength was observed for peripheral and central stimuli. This indicates that task demand is a stronger determinant than input properties, which is consistent with the profound influence of task demands on low-level integration reported in an earlier fMRI study [Van Atteveldt et al., 2007b]. Apparently, speeded detection demands recruit low-level sensory cortices for MSI independent of the eccentricity of the input, possibly to promote fast detection of relevant information both in central as well as peripheral fields. This is also suggested by several studies that found improved detection accompanied by early multisensory interactions for *centrally* presented information [Cappe et al., 2009b, 2012; Romei et al., 2009, in press]. However, it should be noted that the cumulative probabilities of the reaction times (Fig. 2B and Supporting Information Fig. S1b) do indicate stronger multisensory effects for the peripheral inputs during the speed task.

Multisensory Integration in Auditory Versus Visual Cortices

Interestingly, MSI was observed in opposite directions across cortices during the speed task, with enhancement present in the auditory cortex compared with suppression in the visual cortex. The behavioral results as well as the dynamics of the blood oxygen level dependent (BOLD) response in both modalities provide important cues for explaining this different direction of MSI effects. Task performance indicated that the visual information is clearer and more reliable than the sounds, and might therefore be the dominant information for speeded detection in the audiovisual trials, although still susceptible to auditory input as indicated by the AV performance advantage.

Auditory cortical latencies are shorter than visual latencies [Musacchia and Schroeder, 2009; Wang et al., 2008], so the onset of the faster auditory cortical input may speed up visual detection. It has been shown before that auditory stimuli can enhance visual cortex excitability [Romei et al., 2007] and that this can induce earlier spiking responses in visual cortex [Wang et al., 2008]. Speeding up of processing could result in a more transient neural activation and thus a lower and shorter BOLD response, as this is a temporal summation of activity. The dynamics of the reconstructed BOLD responses (Fig. 5B) in the visual ROIs support this interpretation, as the audiovisual response decays faster than to the visual one. Speeding up of neural processing as a mechanism for audiovisual integration in humans has been reported before [Fuhrmann Alpert et al., 2008; Martuzzi et al., 2007; Stekelenburg and Vroomen, 2007; van Wassenhove et al., 2005] and seems most related to the predictive value of the earliest modality [Vroomen and Stekelenburg, 2010]. In the audiovisual trials in the present study, early auditory cortical responses could cue the otherwise temporally unpredictable visual input and thereby speed up visual processing.

In auditory cortex, multisensory information enhances the fMRI response. As the behavioral performance indicates that auditory input is less reliable, the visual information in auditory cortex, rather than speeding up processing, is likely to enhance representations. This interpretation is consistent with findings that multisensory enhancement was specific to the sensory cortex of the non-dominant modality [Giard and Peronnet, 1999]. The enhanced fMRI signal in auditory cortex is furthermore predicted by the inverse effectiveness principle—i.e., MSI is stronger if unisensory inputs are less effective [Stein and Meredith, 1993], which obtains for sounds at the behavioral level in our findings. BOLD signal dynamics (Fig. 5B) again support this interpretation, as the response in the auditory cortex to audiovisual stimuli differs from the response to auditory stimuli in amplitude rather than duration. This explanation may partly relate to the effects of acoustic scanner noise in degrading sound stimuli [Raij et al., 2010], although the different direction of MSI (in auditory vs. visual cortices) was observed exclusively during the speed task, whereas scanner noise was also present during the accuracy task. Furthermore, behavioral results acquired outside the scanner (Supporting Information Fig. S1c) show that auditory task performance was also worse than visual when no acoustic scanner noise was present.

Individual Variation in Multisensory Integration

The suboptimal matching of peripheral versus central behavioral visual performance (Fig. 2C) was only present in a subset of participants (Supporting Information Fig. S1d, “non-matched” subgroup), who showed higher RTs for central pictures relative to peripheral pictures.

Excess eye movements to the peripheral trials were ruled out as explanation (no effect of subgroup on saccade count, Supporting Information Fig. S1d). This suggests that the size difference between peripherally and centrally presented pictures may have been overcompensated in this subgroup, perhaps related to individual differences in cortical magnification values [Dougherty et al., 2003]. Subgroups fMRI analyses revealed a higher-level (non-retinotopic) visual cluster (Supporting Information Fig. S2a), that was specifically involved in speeded detection, and appears to integrate centrally presented AV information best in the subgroup that performed optimally for the central pictures (“matched” subgroup). Furthermore, different MSI profiles for the subgroups in low-level visual cortex were revealed using the individually defined retinotopic ROIs (Supporting Information Fig. S2b). Subgroup differences in these ROIs were found specifically for the accuracy task, and again appear most pronounced for the central stimuli. Together, these findings speak against overcompensation for cortical magnification, because if the central pictures were too small for the “non-matched” subgroup, inverse effectiveness would actually predict stronger rather than weaker MSI. Therefore, the response profiles of these regions indicate individual behavioral skills and/or strategies as factor influencing MSI in sensory cortex [see also Giard and Peronnet, 1999; Romei et al., in press]. For example, the “non-matched” subgroup might have expressed increased attention levels for the peripheral stimuli as compared to the central ones, explaining both the behavioral central disadvantage and weaker MSI for central stimuli.

fMRI Indices of Multisensory Integration

Because thousands of neurons are sampled together in one voxel at typical fMRI resolution, fMRI cannot distinguish between neuronal or regional integration. In the latter case, a stronger response for audiovisual versus unisensory stimulation at the voxel-level is produced by activation of different unisensory neuronal subpopulations within voxels [Laurienti et al., 2005]. An alternative criterion to the one we used, the super-additivity metric, is potentially less sensitive to generate such “false positives” [Calvert, 2001], but at the same time is more sensitive to regional variation in unisensory response strength and BOLD saturation [Goebel and van Atteveldt, 2009]. Moreover, it is actually also affected by the coarse resolution of fMRI voxels as multisensory neurons of different response types (additive, super-additive, sub-additive; [Perrault et al., 2005]) are most probably intermixed within standard-resolution voxels [Beauchamp et al., 2004b; Dahl et al., 2009]. This further decreases the likelihood of finding super-additivity at the voxel-level.

As argued before [Beauchamp, 2005; Doehrmann and Naumer, 2008; Goebel and van Atteveldt, 2009], the employed “max criterion” is a feasible criterion to use in

whole-brain analyses, especially in our case where sensory-specific as well as heteromodal areas are included in the predicted integration circuits. Our results confirm the importance of considering unisensory response profiles as unisensory activation clearly depends on the task (auditory cortex, Fig. 3A), rendering a super-additivity analysis not well comparable across the different conditions. The max criterion has been successfully employed in previous fMRI studies [Hein et al., 2007; Naumer et al., 2008; Van Atteveldt et al., 2004, 2007a]. Moreover, the behavioral results (Millers Inequality test) support the interpretation of the fMRI results as reflecting multisensory integration. Nonetheless, it should be kept in mind that our results do not give conclusive answers about MSI at the level of single neurons.

Conclusion and Future Directions

Our findings demonstrate that MSI within low-level sensory cortex is strongly controlled by stimulus and task context. We also found task-specific MSI in a higher-order processing region (posterior STS/angular gyrus). These findings partly support the predicted functionally specialization for speeded detection in peripheral locations versus discrimination accuracy in foveal locations, but more importantly, they point to a much more flexible use of the available neural architecture for low-level multisensory integration than previously appreciated. The deviations from a clear double dissociation for peripheral/speed and central/accuracy can be summarized as follows. (1) In both low-level auditory and visual cortices, task demand seemed a stronger determinant of MSI, as the necessity for speed overruled any eccentricity effects. When accuracy in identification was required, the effect of eccentricity was actually opposite to our prediction, namely stronger for peripheral stimuli. (2) When speed was required, auditory and visual cortices expressed different MSI processes, namely suppression in visual and enhancement in auditory cortex. We relate these qualitatively different integration processes to timing characteristics for auditory (faster) and visual (slower) first cortical responses, enabling sounds to speed up detection in visual cortex, resulting in a net suppressed fMRI response. (3) Individual variation in behavioral peripheral versus central discrimination abilities or strategies emerges as additional factor influencing MSI in low-level sensory cortex. Future investigations examining context effects on neurofunctional connectivity will help to link our localized findings to specific neural pathways.

ACKNOWLEDGMENTS

The authors thank the developers of the Multimodal Stimulus Set, T.R. Schneider, S. Debener, and A.K. Engel (Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf,

Germany), and Barbara van Doorn, David Semanek, Ben Gunter, and Yunsuo Duan for their assistance with data acquisition.

REFERENCES

- Amedi A, von Kriegstein K, Van Atteveldt NM, Beauchamp MS, Naumer MJ (2005): Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166: 559–571.
- Barnett-Cowan M, Harris L (2009): Perceived timing of vestibular stimulation relative to touch, light and sound. *Exp Brain Res* 198:221–231.
- Bavelier D, Dye MWG, Hauser PC (2006): Do deaf individuals see better? *Trends Cogn Sci* 10:512–518.
- Beauchamp M (2005): Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113.
- Beauchamp M, Lee K, Argall B, Martin A (2004a): Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Beauchamp MS, Argall BD, Bodurka J, Duyn J, Martin A (2004b): Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat Neurosci* 7: 1190–1192.
- 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.
- Bolognini N, Senna I, Maravita A, Pascual-Leone A, Merabet LB (2010): Auditory enhancement of visual phosphene perception: The effect of temporal and spatial factors and of stimulus intensity. *Neurosci Lett* 477:109–114.
- Budinger E, Heil P, Hess A, Scheich H (2006): Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience* 143:1065–1083.
- Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Cappe C, Barone P (2005): Heteromodal connections supporting multisensory integration at low levels of cortical processing. *Eur J Neurosci* 22:2886–2901.
- Cappe C, Morel A, Barone P, Rouiller EM (2009a): The thalamocortical projection systems in primate: An anatomical support for multisensory and sensorimotor interplay. *Cereb Cortex* 19: 2025–2037.
- Cappe C, Morel A, Rouiller EM (2007): Thalamocortical and the dual pattern of corticothalamic projections of the posterior parietal cortex in macaque monkeys. *Neuroscience* 146: 1371–1387.
- Cappe C, Thelen A, Romei V, Thut G, Murray MM (2012): Looming signals reveal synergistic principles of multisensory integration. *J Neurosci* 32:1171–1182.
- Cappe C, Thut G, Romei V, Murray MM (2009b): Selective integration of auditory-visual looming cues by humans. *Neuropsychologia* 47:1045–1052.
- Cappe C, Thut G, Romei V, Murray MM (2010): Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *J Neurosci* 30:12572–12580.
- Dahl CD, Logothetis NK, Kayser C (2009): Spatial organization of multisensory responses in temporal association cortex. *J Neurosci* 29:11924–11932.
- Doehrmann O, Naumer MJ (2008): Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Res* 1242:136–150.
- Dougherty R, Koch V, Brewer A, Fischer B, Modersitzki J, Wandell B (2003): Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J Vis* 3:586–598.
- Driver J, Noesselt T (2008): Multisensory interplay reveals cross-modal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57:11–23.
- Falchier A, Clavagnier S, Barone P, Kennedy H (2002): Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749–5759.
- Falchier A, Schroeder CE, Hackett TA, Lakatos P, Nascimento-Silva S, Ulbert I, Karmos G, Smiley JF (2010): Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb Cortex* 20:1529–1538.
- Forman SD, Cohen JD, Fitzgerald M, Eddy W, Mintun M, Noll DC (1995): Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
- Fort A, Delpuech C, Pernier J, Giard MH (2002): Early auditory-visual interactions in human cortex during nonredundant target identification. *Brain Res Cogn Brain Res* 14:20–30.
- Fuhrmann Alpert G, Hein G, Tsai N, Naumer M, Knight R (2008): Temporal characteristics of audiovisual information processing. *J Neurosci* 28:5344–5349.
- Ghazanfar AA, Schroeder C (2006): Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285.
- Giard MH, Peronnet F (1999): Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490.
- Goebel R, Esposito F, Formisano E (2006): Analysis of Functional Image Analysis Contest (FIAC) data with BrainVoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392–402.
- Goebel R, van Atteveldt N (2009): Multisensory functional magnetic resonance imaging: A future perspective. *Exp Brain Res* 198:153–164.
- Hackett T, Preuss T, Kaas J (2001): Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197–222.
- Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007): Multisensory convergence in auditory cortex. II. Thalamocortical connections of the caudal superior temporal plane. *J Comp Neurol* 502:924–952.
- Hein G, Doehrmann O, Müller NG, Kaiser J, Muckli L, Naumer MJ (2007): Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881–7887.
- Hein G, Knight R (2008): Superior temporal sulcus—It’s my area: Or is it? *J Cogn Neurosci* 20:2125–2136.
- Hocking J, Price CJ (2008): The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex* 18: 2439–2449.
- Kayser C, Logothetis NK (2007): Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212:121–132.
- Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D (2009): Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia* 47:3059–3066.

- 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.
- Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004): Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res* 158:405–414.
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005): On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289–297.
- Lee H, Truy E, Mamou G, Sappey-Marini D, Giraud A (2007): Visual speech circuits in profound acquired deafness: A possible role for latent multimodal connectivity. *Brain* 130:2929–2941.
- Martuzzi R, Murray M, Michel C, Thiran J, Maeder P, Clarke S, Meuli R (2007): Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb Cortex* 17:1672–1679.
- Meinenbrock A, Naumer M, Doehrmann O, Singer W, Muckli L (2007): Retinotopic effects during spatial audio-visual integration. *Neuropsychologia* 45:531–539.
- Miller J (1982): Divided attention: Evidence for coactivation with redundant signals. *Cogn Psychol* 14:247–279.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002): Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cogn Brain Res* 14:115–128.
- Murray MM, Foxe JJ, Higgins BA, Javitt DC, Schroeder CE (2001): Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: A high-density electrical mapping study. *Neuropsychologia* 39:828–844.
- Musacchia G, Schroeder C (2009): Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hear Res* 258:72–79.
- Naumer MJ, Doehrmann O, Mueller NG, Muckli L, Kaiser J, Hein G (2008): Cortical plasticity of audio-visual object representations. *Cereb Cortex* 19:1641–1653.
- Neville H, Lawson D (1987): Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Res* 405:268–283.
- Ochiai T, Grimault S, Scavarda D, Roch G, Hori T, Rivière D, Mangin J, Régis J (2004): Sulcal pattern and morphology of the superior temporal sulcus. *Neuroimage* 22:706–719.
- Perrault TJJ, Vaughan JW, Stein BE, Wallace MT (2005): Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J Neurophysiol* 93:2575–2586.
- Raij T, Ahveninen J, Lin F, Witzel T, Jääskeläinen I, Letham B, Israeli E, Sahyoun C, Vasios C, Stufflebeam S, Hämäläinen M, Belliveau J (2010): Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci* 31:1772–1782.
- Rockland KS, Ojima H (2003): Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol* 50:19–26.
- Röder B, Teder-Sälejärvi W, Sterr A, Rösler F, Hillyard S, Neville H (1999): Improved auditory spatial tuning in blind humans. *Nature* 400:162–166.
- Romei V, Murray M, Cappe C, Thut G (2009): Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Curr Biol* 19:1799–1805.
- Romei V, Murray M, Cappe C, Thut G: The contributions of sensory dominance and attentional bias to cross-modal enhancement of visual cortex excitability. *J Cogn Neurosci* (in press).
- Romei V, Murray MM, Merabet LB, Thut G (2007): Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *J Neurosci* 27:11465–11472.
- Schneider T, Engel A, Debener S (2008): Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp Psychol* 55:121–132.
- Schroeder CE, Molholm S, Lakatos P, Ritter W, Foxe J (2004): Human-simian correspondence in the early cortical processing of multisensory cues. *Cogn Process* 5:140–151.
- Schroeder CE, Smiley JF, Fu KG, McGinnis T, O'Connell MN, Hackett TA (2003): Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int J Psychophysiol* 50:5–17.
- Shams L, Kamitani Y, Thompson S, Shimojo S (2001): Sound alters visual evoked potentials in humans. *Exp Brain Res* 12:3849–3852.
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE (2007): Multisensory convergence in auditory cortex I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502:894–923.
- Sperdin HF, Cappe C, Foxe JJ, Murray MM (2009): Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front Integr Neurosci* 3:1–10.
- Stein BE, Meredith MA (1993): *The Merging of the Senses*. Cambridge, MA: MIT Press.
- Stekelenburg J, Vroomen J (2007): Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964–1973.
- Stevenson R, Kim S, James T (2009): An additive-factors design to disambiguate neuronal and areal convergence: Measuring multisensory interactions between audio, visual, and haptic sensory streams using fMRI. *Exp Brain Res* 198:183–194.
- Takeshima Y, Gyoba J (2013): High-intensity sound increases the size of visually perceived objects. *Atten Percept Psychophys* 75:501–507.
- Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006): Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci USA* 102:8239–8244.
- Van Atteveldt N, Formisano E, Goebel R, Blomert L (2004): Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- Van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007a): The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962–974.
- Van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007b): Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36:1345–1360.
- Van Atteveldt NM, Roebroek A, Goebel R (2009): Interaction of speech and script in human auditory cortex: Insights from neuro-imaging and effective connectivity. *Hear Res* 258:152–164.
- van Wassenhove V, Grant K, Poeppel D (2005): Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci* 102:1181–1186.
- von Kriegstein K, Griffiths T, Thompson S, McAlpine D (2008): Responses to interaural time delay in human cortex. *J Neurophysiol* 100:2712–2718.

- Vroomen J, Stekelenburg J (2010): Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J Cogn Neurosci* 22:1583–1596.
- Wang Y, Celebrini S, Trotter Y, Barone P (2008): Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence. *BMC Neurosci* 9:79.
- Welford AT (1980): Choice reaction time: Basic concepts. In: Welford AT, editor. *Reaction Times*. New York: Academic Press. pp 73–128.
- Werner S, Noppeney U (2010): Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *J Neurosci* 30:2662–2675.
- Woodworth RS, Schlosberg H (1954): *Experimental Psychology*. New York: Henry Holt.
- Zhang N, Chen W (2006): A dynamic fMRI study of illusory double-flash effect on human visual cortex. *Exp Brain Res* 172: 57–66.