Behavioral/Systems/Cognitive

Object Familiarity and Semantic Congruency Modulate Responses in Cortical Audiovisual Integration Areas

Grit Hein,[1,2,4]* **Oliver Doehrmann,**[3] **Notger G. Müller,**[1,2] **Jochen Kaiser,**[3] **Lars Muckli,**[2,5] **and Marcus J. Naumer**[2,3,5]*

[1]Cognitive Neurology Unit, [2]Brain Imaging Center, and [3]Institute of Medical Psychology, Johann Wolfgang Goethe-University, D-60528 Frankfurt am Main, Germany, [4]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, and [5]Department of Neurophysiology, Max Planck Institute for Brain Research, D-60528 Frankfurt am Main, Germany

The cortical integration of auditory and visual features is crucial for efficient object recognition. Previous studies have shown that audiovisual (AV) integration is affected by where and when auditory and visual features occur. However, because relatively little is known about the impact of what is integrated, we here investigated the impact of semantic congruency and object familiarity on the neural correlates of AV integration. We used functional magnetic resonance imaging to identify regions involved in the integration of both (congruent and incongruent) familiar animal sounds and images and of arbitrary combinations of unfamiliar artificial sounds and object images. Unfamiliar object images and sounds were integrated in the inferior frontal cortex (IFC), possibly reflecting learning of novel AV associations. Integration of familiar, but semantically incongruent combinations also correlated with IFC activation and additionally involved the posterior superior temporal sulcus (pSTS). For highly familiar semantically congruent AV pairings, we again found AV integration effects in pSTS and additionally in superior temporal gyrus. These findings demonstrate that the neural correlates of object-related AV integration reflect both semantic congruency and familiarity of the integrated sounds and images.

Key words: multisensory integration; cross-modal; object recognition; familiarity; semantic congruency; functional magnetic resonance imaging; fMRI

Introduction

The integration of auditory and visual object features is a crucial aspect of efficient object recognition. Single-cell studies (Meredith and Stein, 1983, 1996) and more recent functional magnetic resonance imaging (fMRI) experiments (for review, see Calvert, 2001; Amedi et al., 2005) have demonstrated that neural responses to audiovisual (AV) stimulation are most pronounced for stimuli which coincide in space and time. In contrast, as yet relatively little is known about the potential impact of semantic AV object features on the topography and strength of cortical responses.

So far, only very few studies included an explicit manipulation of semantic AV congruency. Laurienti et al. (2004) have shown that semantically congruent, but not incongruent AV combinations result in a behavioral performance enhancement. On the neural level, integration of congruent AV combinations evoked stronger activations of higher-level visual (Belardinelli et al.,

2004) and auditory regions (van Atteveldt et al., 2004, 2007) and superior temporal sulcus (STS) (Calvert et al., 2000). Posterior temporal regions around STS (pSTS) and middle temporal gyrus (MTG) appeared to be slightly more involved in integration of semantically congruent than incongruent combinations (Beauchamp et al., 2004; van Atteveldt et al., 2004), but were also significantly activated by incongruent AV pairings (Taylor et al., 2006). In line with these human neuroimaging studies, there is also electrophysiological evidence for AV integration of congruent sounds and images in primate auditory cortex (Ghanzanfar et al., 2005) and anterior STS (Barraclough et al., 2005). Moreover, single-cell data indicate an involvement of primate prefrontal cortex in the integration of both abstract AV material, such as colors and tones (Fuster et al., 2000) and natural communication signals (Sugihara et al., 2006). Supporting these findings, stronger AV integration effects for semantically incongruent than congruent AV object features have also been reported in human fMRI studies (Belardinelli et al., 2004; Taylor et al., 2006).

Together, these previous studies suggest a number of cortical candidate regions in human temporal and frontal lobes for object-related AV integration. However, so far, fMRI studies have only used highly familiar sounds and images (Beauchamp et al., 2004a,b; Belardinelli et al., 2004; van Atteveldt et al., 2004; 2007; Taylor et al., 2006). This allowed the variation of semantic congruency, whereas the impact of stimulus familiarity on the neural correlates of object-related AV integration has remained mostly unknown.

In the present fMRI study, we investigated the impact of both

semantic congruency and stimulus familiarity on the aforementioned frontotemporal cortical network. We compared the integration of unfamiliar artificial object images ("fribbles") and sounds to AV integration of highly familiar animal images and sounds that were presented either in semantically congruent (e.g., dog image and barking sound) or incongruent (e.g., dog image and meowing sound) pairings. We were able to reveal a frontotemporal network for object-related AV integration whose several regions show differential sensitivities to semantic congruency and to stimulus familiarity.

Materials and Methods

AV main experiment

Participants. Eighteen subjects (seven female; mean age, 29.8; range, 23–41 years, one left handed) participated in the study. All subjects had normal or corrected-to-normal vision and hearing. All participants received information on MRI and a questionnaire to check for potential health risks and contraindications. Volunteers gave their informed consent after having been introduced to the procedure in accordance with the declaration of Helsinki.

Stimuli. In the familiar conditions, we used eight different animal sounds and images. The unfamiliar material comprised eight images of artificial objects (fribbles; http://α.cog.brown.edu:8200/stimuli/novel-objects/fribbles.zip/view) (see Fig. 1*a*) and eight artificial sounds created by the distortion (played backwards and with underwater effect) of the animal sounds used in the familiar conditions. In a behavioral pretest, artificial sounds were played to eight subjects and none of these sounds was associated with any familiar object.

The stimuli were presented using Neurobehavioral Systems (Albany, CA) presentation software running on a personal computer (Miditower Celeron) at a frame rate of 60 Hz. Sounds and images were presented in stimulation blocks at a rate of one every 2000 ms. Stimulation blocks consisted of eight stimulus events with a fixation cross being constantly present. Images were projected onto a vertical screen positioned inside the scanner with a liquid crystal display projector (VPL PX 20; Sony, Tokyo, Japan) equipped with a custom-made lens. Subjects viewed the screen through a mirror. The mirror and projection screen were fixed on the head coil. The subjects' field of view was 52.5° visual angle (maximum horizontal distance). Visual stimulation consisted of gray-scaled photographs (mean stimulus size, 14.6° angle), which were presented in the center of a black screen.

Auditory stimuli were presented through an MRI audio system (Commander XG, Resonance Technology, Northridge, CA) (frequency response, 100 to ± 25 kHz). Subjects received them via headphones simultaneously to both ears. Spectrograms of three representative sounds are shown in Figure 1. Images and sounds are available on request from the authors.

Procedure. Animal sounds, animal images, artificial sounds, and fribble images were presented in eight experimental conditions [unimodal auditory animal/artificial, unimodal visual animal/fribble, AV familiar congruent (e.g., dog-barking), AV familiar incongruent (e.g., dog-meowing), and AV unfamiliar artificial in fixed order, and AV unfamiliar artificial in random order]. During bimodal conditions, sounds and images were presented simultaneously.

Each of these eight experimental conditions was presented for ∼16 s (eight measurement volumes) in a block design, separated from the next block by a fixation period of equal length. A complete experimental run comprised two cycles of experimental conditions plus an additional eight volumes of fixation at the beginning of the run (280 volumes). The session had five experimental runs, including all experimental conditions in randomized order. Subjects were asked to fixate and be attentive during the measurements. We have chosen a passive paradigm to minimize task-related frontal activation (Calvert et al., 2000; Belardinelli et al., 2004; van Atteveldt et al., 2004, 2007). Due to the sluggishness of the blood oxygen level-dependent (BOLD) signal, it is otherwise hard to compellingly disentangle task-related frontal activation from frontal involvement in AV integration.

Imaging. fMRI was performed on a 3 Tesla Siemens (Erlangen, Germany) Magnetom Allegra scanner at the Brain Imaging Center in Frankfurt am Main. A gradient-recalled echo-planar imaging sequence was used with the following parameters: 34 slices; repetition time (TR), 2000 ms; echo time (TE), 30 ms; field of view, 192 mm; in-plane resolution, 3×3 mm^2; slice thickness, 3 mm; gap thickness, 0.3 mm. For each subject, a magnetization-prepared rapid-acquisition gradient-echo sequence was used (TR, 2300 ms; TE, 3.49 ms; flip angle, 12°; matrix, 256×256; voxel size, $1.0 \times 1.0 \times 1.0$ mm^3) for detailed anatomical imaging.

Data analysis. Neural correlates of AV integration were assessed separately for each of the four bimodal conditions. We thereby searched for regions that were (1) significantly activated during each of the unimodal conditions [audio (A); visual (V)], and (2) responded more strongly to bimodal AV stimulation than to each of the unimodal conditions. Accordingly, the identification of brain regions involved in object-related AV integration was based on significant activation in a $[(AV > A) \cap (AV > V) \cap (A > 0) \cap (V > 0)]$ conjunction analysis. In a previous debate, these criteria have been suggested as sufficiently strict and appropriate for the definition of multisensory integration effects in human brain imaging studies (Beauchamp et al., 2004a; Beauchamp, 2005a; Laurienti et al., 2005) (but see Calvert et al., 2001).

Data were analyzed using the BrainVoyager QX (Brain Innovation, Maastricht, the Netherlands) software package. The first four volumes of each experimental run were discarded to preclude T1 saturation effects. Preprocessing of the functional data included the following steps: (1) three-dimensional motion correction, (2) linear-trend removal and temporal high-pass filtering at 0.0054 Hz, (3) slice-scan-time correction with sinc interpolation, and (4) spatial smoothing using a Gaussian filter of 8 mm (full width at half maximum). Volume-based statistical analyses were performed using a random effects general linear model (df = 17). For every voxel, the time course was regressed on a set of dummy-coded predictors representing the experimental conditions. To account for the shape and delay of the hemodynamic response (Boynton et al., 1996), the predictor time courses (box-car functions) were convolved with a gamma function. Statistical maps were corrected for multiple comparisons using cluster-size thresholding (Forman et al., 1995; Goebel et al., 2006). In this method, for each statistical map the uncorrected voxel-level threshold was set at $t > 3$ ($p < 0.009$; unless otherwise indicated in the respective figure legends) and was then submitted to a whole-brain correction criterion based on the estimate of the spatial smoothness of the map and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates. After 1,000 iterations, the minimum cluster-size that yielded a cluster-level false-positive rate of 5% was used to threshold the statistical map.

Group-averaged functional data were then projected on inflated representations of the left and right cerebral hemispheres of one subject. Because a morphed surface always possesses a link to the folded reference mesh, functional data can be shown at the correct location of folded as well as inflated representations. This link was also used to keep geometric distortions to a minimum during inflation through inclusion of a morphing force that keeps the distances between vertices and the area of each triangle of the morphed surface as close as possible to the respective values of the folded reference mesh.

Visuotactile control experiment

In addition, we conducted a visuotactile (VT) control experiment to test whether the effects revealed in the main experiment reflected increased complexity of stimulation during bimodal compared with unimodal stimulation instead of AV integration.

Participants. We recorded fMRI in 11 subjects, who had also participated in our AV main experiment.

Stimuli and procedure. Subjects viewed and/or touched unfamiliar artificial objects (wooden fribbles) and familiar objects (toy animals). All objects were presented in both unimodal (visual and tactile) as well as bimodal (VT) conditions. Familiar VT combinations consisted of touchable animals and their respective photographs that were presented either canonically (VT congruent) or horizontally mirrored (VT incongruent). Using a block design, each experimental condition was presented for ∼20 s, (10 measurement volumes; TR, 2 s) separated from the next block by a fixation period of equal length. Each single stimulus was presented

a **unfamiliar artificial**

„pwfoosh"

b **familiar incongruent**

„meow"

c **familiar congruent**

„woof-
woof"
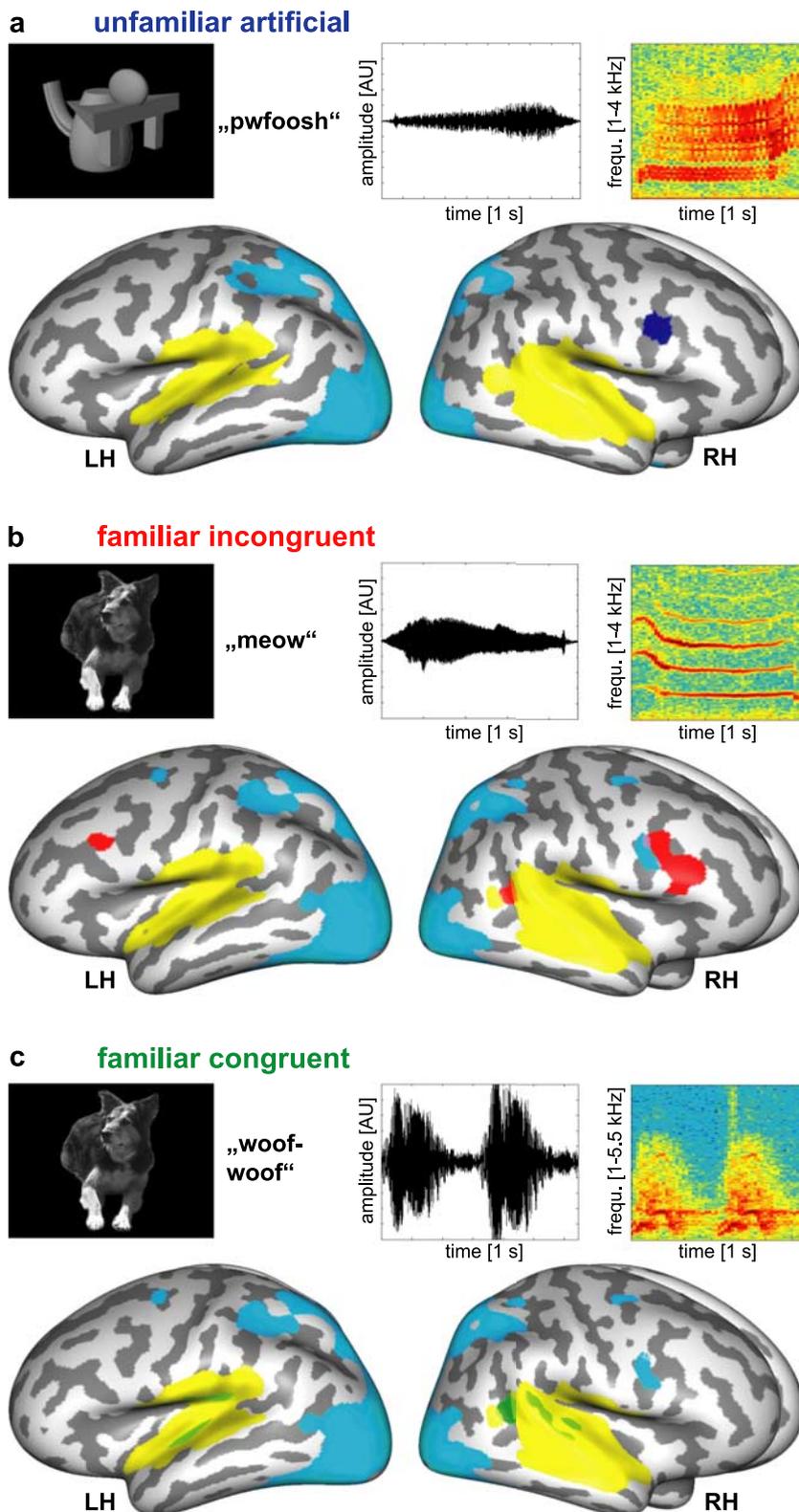
Figure 1. *a–c,* Visual and auditory (amplitude waveforms and spectrograms) example stimuli and cortical regions of object-related AV integration [defining contrast, $0 < A < AV > V > 0$; $t > 3.0$; p (corrected) < 0.05] for unfamiliar artificial (*a*) familiar incongruent (*b*), and familiar congruent conditions (*c*). Unimodal auditory (yellow) and visual (light blue) activations [q(false discovery rate) < 0.01] are shown. LH, Left hemisphere; RH, right hemisphere; AU, arbitrary units.

for 2 s and reopening of the right hand was cued by a color change of the fixation cross. A complete experimental run comprised two cycles of experimental conditions plus an additional 10 volumes of fixation at the beginning of each run (350 volumes). The session consisted of four ex-

perimental runs each including all experimental conditions in randomized order.

Imaging. The acquisition of anatomical and functional images and the preprocessing of the functional imaging data were identical to that in our AV main experiment.

Data analysis. Activations in unimodal and bimodal conditions of the VT control experiment were analyzed region-of-interest (ROI)-based in the same frontal and temporal regions which were identified as AV integration sites in the AV main experiment.

Results
AV main experiment
Unimodal activations
Both unimodal contrasts (A > V; V > A) revealed activations of the respective modality-specific cortices. Figure 1 shows the pattern of activations during unimodal visual (light blue) and auditory stimulations (yellow). Strength and distribution of activation was comparable for familiar animal sounds (Fig. 1*b,c*) and unfamiliar artificial sounds (Fig. 1*a*). The same was true for familiar animal images (Fig. 1*b,c*) and unfamiliar fribble images (Fig. 1*a*).

Bimodal activations
Neural correlates of object-related AV integration for unfamiliar artificial, familiar incongruent, and familiar congruent material are depicted in Figures 1*a–c* (indicated by blue, red, and green, respectively). Because no significant differences were found between fixed and random combinations of unfamiliar artificial images and sounds, we pooled the data from these two experimental conditions.

AV integration effects for unfamiliar artificial object features are shown in Figure 1*a*. A conjunction analysis comparing the respective bimodal and unimodal conditions [$0 < A < AV > V > 0$; p (corrected) < 0.05] revealed an activation in right inferior frontal cortex (IFC) (Fig. 1*a*, dark blue). AV integration effects for familiar incongruent object features are depicted in Figure 1*b* [red; $0 < A < AV > V > 0$; p (corrected) < 0.05]. Again, we revealed substantial IFC activations. In the right hemisphere they extended ventrally along the inferior frontal gyrus (IFG). Apart from these frontal effects, AV integration of familiar incongruent features also activated right pSTS. AV integration of familiar congruent object features correlated with activation in right pSTS and bilateral superior temporal gyrus (STG) (Fig. 1*c*, green) [$0 < A < AV > V > 0$; p (corrected) < 0.05]. Figure 1*c* shows a substantial overlap between this AV integration effect (green) and the auditory preference map (yellow).

Together, our results showed overlapping AV integration effects for unfamiliar artificial and familiar incongruent object fea-
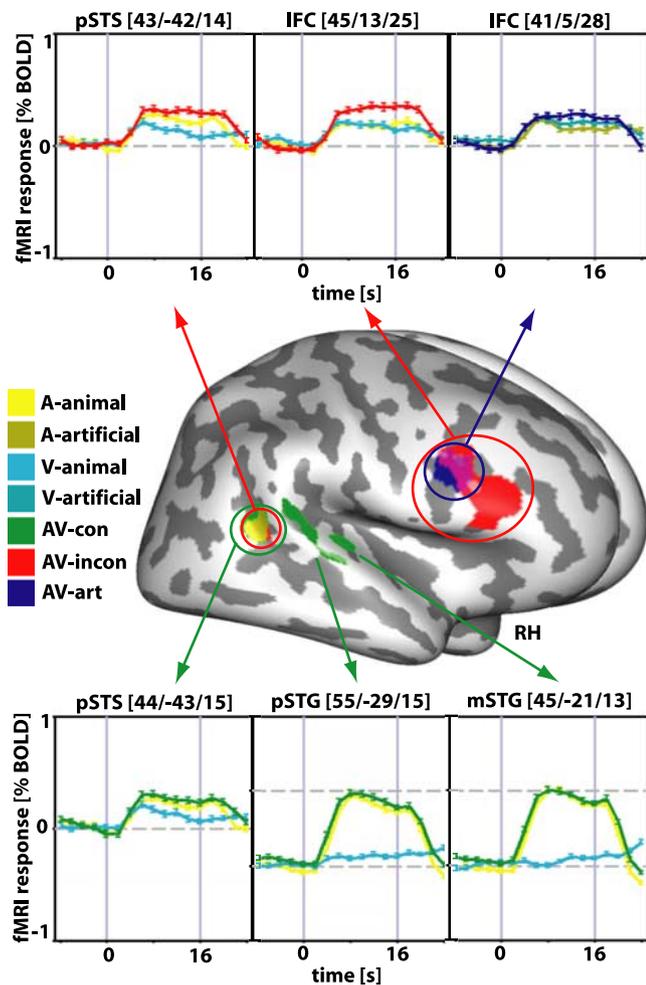
Figure 2. Overlap of significantly activated AV integration regions [defining contrast, 0 < A < AV > V > 0; t >3.0; p (corrected) < 0.05] for unfamiliar artificial (blue), familiar incongruent (red), and familiar congruent object features (green), and the average percentage of BOLD signal change during bimodal and unimodal conditions. RH, Right hemisphere; p, posterior; m, mid.

tures in right IFC (purple; Fig. 2). Integration of familiar incongruent, but not unfamiliar artificial AV combinations further elicited an activation of ventral IFC along IFG (red). AV integration effects in pSTS for familiar incongruent and familiar congruent object features also overlapped (yellow). Finally, STG activation correlated with integration of highly familiar congruent material (green).

We also computed direct statistical contrasts between the familiar animal and unfamiliar artificial versus familiar incongruent conditions (blue). Figure 3 shows significant activations of the contrasts between familiar congruent versus unfamiliar artificial (green), familiar incongruent versus unfamiliar artificial (red), and unfamiliar artificial versus familiar incongruent conditions (blue). Contrasting familiar congruent (con) and unfamiliar artificial (art) AV combinations [AV_con > AV_art ∩ AV_con >0; t = 4; p (corrected) < 0.05] revealed extended activations in bilateral STG, pSTS, and MTG regions, predominantly including regions of higher-level auditory processing (green) (Lewis et al., 2005). These activations largely overlapped with those for the contrast between familiar incongruent (incon) and unfamiliar artificial material [AV_incon > AV_art ∩ AV_incon > 0; t = 4; p (corrected) < 0.05; red and yellow]. In addition, familiar incongruent AV combinations led to stronger activa-

tions along the bilateral IFG and in the occipital cortex (red). The contrast between unfamiliar artificial and familiar incongruent AV conditions [AV_art > AV_incon ∩ AV_art > 0; t = 4; p (corrected) < 0.05] revealed dorsal occipital and posterior parietal activations. These regions might be associated with action-related visual processing (Milner and Goodale, 1995) rather than AV integration, because we did not obtain significant parietal or occipital activations in the [0 < A < AV > V > 0] conjunction analysis (Figs. 1, 2). Contrasting familiar congruent and incongruent AV conditions did not reveal any significant effects.

Visuotactile control experiment

So far, our results show significant integration effects in bilateral IFC, right pSTS, and bilateral STG. However, it could be argued that this might simply reflect the more complex stimulation during bimodal compared with unimodal conditions rather than AV integration. In our VT control experiment, we explicitly tested this alternative explanation with 11 of our 18 subjects. We used the same overall experimental design, but visual and tactile, instead of auditory and visual, stimulation. If the integration effects in the AV main experiment were mainly caused by the higher complexity of bimodal compared with unimodal stimulation, bimodal VT stimulation should also elicit stronger activations in the IFC, pSTS, and STG than unimodal visual and tactile conditions.

Activation elicited by VT stimulation in IFC, pSTS, and STG regions, which were identified as AV integration sites in our main experiment (Fig. 2) are shown in Figure 4. In contrast to our AV main experiment, activations in these ROIs during the VT control experiment were comparable for bimodal and unimodal conditions (Fig. 4).

Discussion

In the present study, we investigated AV integration of semantically congruent and incongruent familiar animal vocalizations and images and arbitrary pairings of unfamiliar artificial sounds and "fribble" images. Our findings indicate that neural correlates of object-related AV integration depend on both semantic congruency and familiarity of the sounds and images. Integration of unfamiliar artificial sounds and images as well as familiar incongruent material involved inferior frontal regions. Integration of familiar sounds and images correlated with pSTS activation, independently of semantic congruency. Moreover, integration of highly familiar congruent AV material activated higher-order auditory regions in the STG.

AV integration effects in frontal, lateral temporal, and superior temporal regions have been reported in previous studies that used only familiar images and sounds. Belardinelli et al. (2004) and Taylor et al. (2006) reported such effects in the IFC, Beauchamp et al. (2004a,b), van Atteveldt et al. (2004, 2007), and Taylor et al. (2006) showed AV integration in the pSTS, and van Atteveldt et al. (2004, 2007) found similar effects during AV letter processing in the human auditory cortex. However, the heterogeneity of these findings across studies has made it hard to specify the particular contributions of the IFC, pSTS, and STG to object-related AV integration. In our study, we were able to reveal AV integration effects in each of these regions and, moreover, could demonstrate that these integration effects depend on both semantic congruency and stimulus familiarity. Thus, the current findings provide additional insight into the interplay of frontal, lateral temporal, and superior temporal regions during AV object recognition.
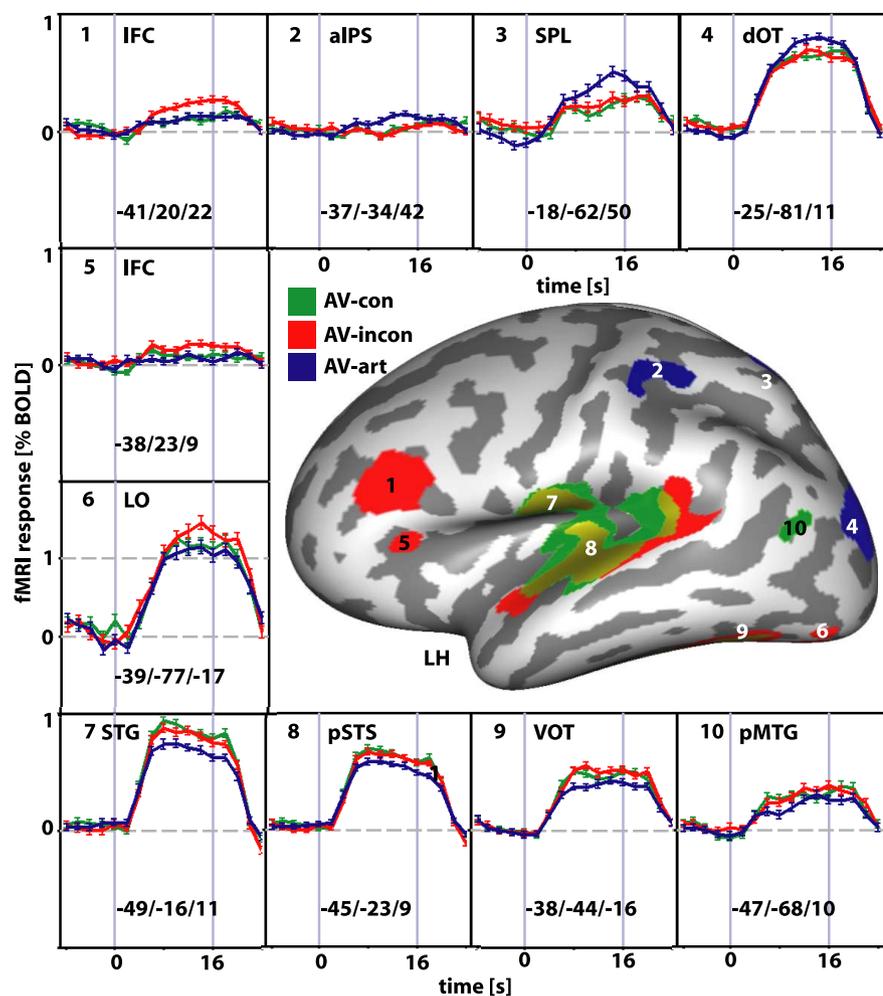
Figure 3. Direct contrasts between bimodal conditions [$t > 4.0$; p (corrected) < 0.05]. Regions activated more strongly by familiar congruent compared with unfamiliar artificial AV object features (green; additional criterion, AV_con > 0), regions activated more strongly by familiar incongruent compared with unfamiliar artificial AV object features (red; additional criterion, AV_incon > 0), and regions activated more strongly by unfamiliar artificial compared with familiar incongruent AV object features (blue; additional criterion, AV_art > 0) are shown. Surrounding panels indicate the average percentage of BOLD signal change in left hemisphere regions with different bimodal response preferences. LH, Left hemisphere; a, anterior; p, posterior; v, ventral; d, dorsal; IPS, intraparietal sulcus; SPL, superior parietal lobule; LO, lateral occipital cortex; OT, occipitotemporal cortex; AV-con, familiar congruent audiovisual stimulation; AV-incon, familiar incongruent audiovisual stimulation; AV-art, unfamiliar artificial audiovisual stimulation.

AV integration of object features in the frontal cortex

Current knowledge about the role of the IFC in AV integration is sparse, because the multifunctionality of frontal regions often makes it difficult to disentangle AV integration from other frontal functions. Particularly critical confounds are frontal activations caused by explicit or even implicit task requirements (Miller and Cohen, 2001) or increasing complexity of stimulation (Duncan and Owen, 2000). Here, we used a passive paradigm, which reduced the impact of task-related effects. Moreover, the results of our VT control experiment allowed us to rule out the assumption that increased activation in the IFC, pSTS, and STG reflected increased stimulus complexity under bimodal compared with unimodal conditions. Based on these data, we are confident that the observed effects are predominantly related to the integration of object sounds and images.

In line with previous studies (Belardinelli et al., 2004; Taylor et al., 2006), we found frontal effects for the integration of familiar incongruent sounds and images, but not for familiar congruent pairings. Extending these studies, we could show that the IFC is also the prime integration site for unfamiliar artificial object images and sounds. Together, these results indicate that AV integration in the IFC reflects semantic congruency, but is rather independent of stimulus familiarity. The substantial frontal activation for familiar incongruent stimuli might partly reflect the detection of semantic violations or bizarre combinations of material (Michelon et al., 2003). However, overlapping with these activations we also found AV integration effects for unfamiliar artificial images and sounds, which were semantically unrelated and therefore unlikely to cause a semantic violation. In the context of these results, AV integration of familiar incongruent and unfamiliar artificial object features in the IFC might reflect both the revision of familiar, and the learning of novel AV associations (Gonzalo et al., 2000). With respect to a possible role of IFC in learning of novel AV associations, stronger AV integration effects in IFC might be expected for random combinations of unfamiliar artificial images and sounds than for fixed pairs. However, our results did not show significant differences in AV integration effects between the fixed and random artificial AV conditions. In our paradigm, each condition was repeated only twice per run, and passively perceived by the subjects. It is possible that particular (fixed) combinations of unfamiliar artificial sounds and images have to be presented more often to establish associations that differentiate them from random pairings.

AV integration of object features in lateral temporal cortex

Our findings are in line with the results of previous studies that have indicated an important role of pSTS in AV integration (Beauchamp et al., 2004a,b; van Atteveldt et al., 2004, 2007; Taylor et al., 2006). Extending these findings, our data allow us to further specify the integrative function of pSTS by demonstrating integration of familiar images and sounds, but not of unfamiliar artificial object features. Thus, AV integration in pSTS reflects stimulus familiarity, but appears to be rather independent of semantic congruency. This is the opposite pattern of what we found for AV integration in the IFC (i.e., sensitivity to semantic congruency and insensitivity to stimulus familiarity), which might imply that the integrative functions of pSTS and IFC complement each other. AV integration in the IFC might serve the learning and re-establishing of associations between AV object features, which, once they have been learned, are integrated in the pSTS. Familiar object sounds are always associated with an image (and vice versa) and, thus, activate pSTS, no matter if they are semantically congruent or incongruent. In semantically incongruent pairings, AV associations have to be revised and potentially re-established, which is reflected in an additional IFC activation. In contrast to familiar incongruent AV object features, there is no
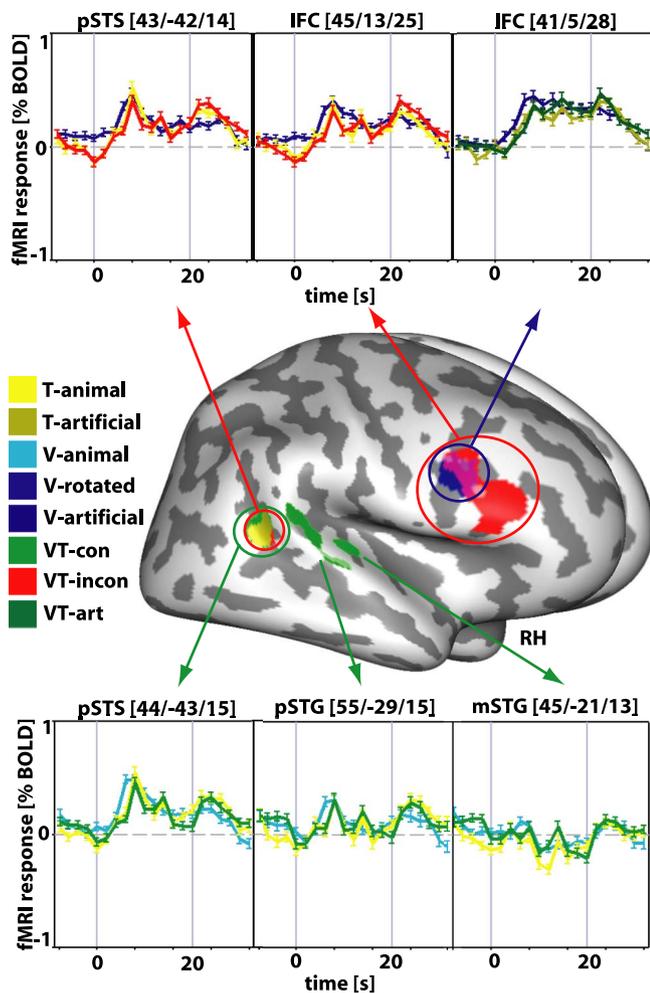
Figure 4. Results of the VT control experiment. The central panel shows the overlap of significantly activated AV integration regions for unfamiliar artificial (blue), familiar incongruent (red), and familiar congruent object features (green; from Fig. 2). Top and bottom panels show the average percentage of BOLD signal change in these regions during unimodal and bimodal conditions of our VT control experiment. In contrast to the results of the AV main experiment, there was no activation increase during bimodal VT stimulation compared with the respective unimodal conditions. Thus, it is unlikely that the observed AV activation increases (main experiment) reflect the higher number of inputs in bimodal compared with unimodal stimulation. RH, Right hemisphere; p, posterior; m, mid; V, unimodal visual stimulation; T, unimodal tactile stimulation; VT, bimodal visuotactile stimulation. art, artificial; con, congruent; incon, incongruent.

need to re-examine familiar congruent AV combinations. They therefore might be suitable for integration in pSTS and even more specialized regions in the STG. In line with this assumption, previous primate data suggest that congruent, but not incongruent or artificial AV combinations are integrated in the anterior STS (Barraclough et al., 2005) and auditory cortex (Ghanzanfar et al., 2005). Our data provide evidence for a role of human auditory cortex in AV integration of highly familiar and semantically congruent object features.

AV integration of object features in superior temporal cortex

For familiar congruent AV combinations, we found AV integration effects in the bilateral STG (Figs. 1, 2). These AV integration sites were located in close vicinity to the mid-STG region, which has been shown to be involved in the processing of animal vocalizations (Lewis et al., 2005). Previous studies have shown AV integration effects for natural or common objects in visual, but

not auditory regions (Belardinelli et al., 2004). Belardinelli et al. (2004) have used a mixture of living and nonliving objects, whereas in our study all meaningful objects belonged to the animal category. As the findings of Lewis et al. (2005) have indicated differences in the neural representation between animal vocalizations and tool sounds, AV objects from different object categories might be integrated in different category-related regions. Our finding of AV integration in the human auditory cortex is in line with previous results by van Atteveldt et al. (2004, 2007), who have reported AV integration effects for congruent combinations of written letters and speech sounds in early auditory regions (Heschl's gyrus). Given Lewis et al.'s (2005) results, it seems plausible that the images and sounds of animals are integrated in higher-level auditory regions, such as those revealed in our study. We found AV integration effects in higher-level auditory regions only for highly familiar and semantically congruent AV material with well established associations between object sounds and images. Based on these findings, it would be interesting to test whether even sensory-specific regions become involved in AV integration of artificial object features if the respective associations are explicitly trained (i.e., if a semantic AV relationship is established and familiarity is substantially increased).

Conclusions

Comparing AV integration of unfamiliar artificial and highly familiar animal stimuli, we were able to demonstrate that cortical AV integration sites in a frontotemporal network are differently activated depending on semantic congruency and stimulus familiarity. AV integration in IFC was found to be sensitive to semantic congruency, but did not depend on stimulus familiarity. In contrast, pSTS integrated AV object features independently of semantic congruency, but was sensitive to stimulus familiarity. Higher-level auditory regions within the STG integrated material that was both highly familiar and semantically congruent. Based on these findings, we propose that frontal and temporal regions might have complementary roles in object-related AV integration.

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.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377–391.

Beauchamp MS (2005) Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3:93–113.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004a) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 4:809–823.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004b) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 11:1190–1192.

Belardinelli MO, Sestieri C, Di Matteo R, Delogu F, Del Gratta C, Ferretti A, Caulo M, Tartaro A, Romani GL (2004) Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. Cogn Process 5:167–174.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.

Calvert GA, Campell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. Curr Biol 10:649–657.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. NeuroImage 14:427–438.

Duncan J, Owen AM (2000) Common regions in the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci 23:475–483.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636–647.

Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal associations in neurons of frontal cortex. Nature 405:347–351.

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

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 27:392–401.

Gonzalo D, Shallice T, Dolan R (2000) Time-dependent changes in learning audiovisual associations: a single-trial fMRI study. NeuroImage 11:243–255.

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.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus of alert cats. Science 221:389–391.

Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. J Neurophysiol 75:1843–1857.

Michelon P, Snyder AZ, Buckner RL, McAvoy M, Zacks JM (2003) Neural correlates of incongruous visual information. An event-related fMRI study. NeuroImage 19:1612–1626.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202.

Milner AD, Goodale MA (1995) The visual brain in action. Oxford: Oxford UP.

Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006) Binding crossmodal object features in perirhinal cortex. Proc Natl Acad Sci USA 21:8239–8244.

Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J Neurosci 25:11138–11147.

van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) 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 (2004) Integration of letters and speech sounds in the human brain. Neuron 43:271–282.