Cerebral Cortex July 2009;19:1641-1653
doi:10.1093/cercor/bhn200
Advance Access publication November 17, 2008

Cortical Plasticity of Audio–Visual Object Representations

Marcus J. Naumer[1], Oliver Doehrmann[1], Notger G. Müller[2,3], Lars Muckli[2,4,5], Jochen Kaiser[1] and Grit Hein[2,3,6]

[1]Institute of Medical Psychology, Goethe-University, D-60528 Frankfurt am Main, Germany, [2]Brain Imaging Center, Goethe-University, D-60528 Frankfurt am Main, Germany, [3]Cognitive Neurology Unit, Goethe-University, D-60528 Frankfurt am Main, Germany, [4]Department of Neurophysiology, Max Planck Institute for Brain Research, D-60528 Frankfurt am Main, Germany, [5]Department of Psychology, University of Glasgow, G12 8QQ Glasgow, UK and [6]Helen Wills Neuroscience Institute, University of California - Berkeley, CA 94720, USA

Marcus J. Naumer, Oliver Doehrmann, and Grit Hein contributed equally to this work.

Several regions in human temporal and frontal cortex are known to integrate visual and auditory object features. The processing of audio–visual (AV) associations in these regions has been found to be modulated by object familiarity. The aim of the present study was to explore training-induced plasticity in human cortical AV integration. We used functional magnetic resonance imaging to analyze the neural correlates of AV integration for unfamiliar artificial object sounds and images in naïve subjects (PRE training) and after a behavioral training session in which subjects acquired associations between some of these sounds and images (POST-training). In the PRE-training session, unfamiliar artificial object sounds and images were mainly integrated in right inferior frontal cortex (IFC). The POST-training results showed extended integration-related IFC activations bilaterally, and a recruitment of additional regions in bilateral superior temporal gyrus/sulcus and intraparietal sulcus. Furthermore, training-induced differential response patterns to mismatching compared with matching (i.e., associated) artificial AV stimuli were most pronounced in left IFC. These effects were accompanied by complementary training-induced congruency effects in right posterior middle temporal gyrus and fusiform gyrus. Together, these findings demonstrate that short-term cross-modal association learning was sufficient to induce plastic changes of both AV integration of object stimuli and mechanisms of AV congruency processing.

Keywords: congruency, cortex, cross-modal, functional magnetic resonance imaging, human, multisensory, object perception

Introduction

In the last decade, audio-visual (AV) integration of object images and sounds has been investigated in numerous studies. AV integration sites have been identified in superior temporal sulcus (STS; Calvert et al. 2001; Beauchamp et al. 2004; Taylor et al. 2006; Hein et al 2007), inferior frontal cortex (IFC; Belardinelli et al. 2004; Taylor et al. 2006; Hein et al. 2007), auditory (superior temporal gyrus, STG; Hein et al. 2007; Heschl's gyrus, HG; van Atteveldt et al. 2004, 2007) and visual cortex (Belardinelli et al. 2004). There is increasing evidence that the identified regions have different AV integration profiles. AV integration in auditory and visual regions has been found mainly for familiar and congruent sounds and images, for example a dog picture and a barking sound (Belardinelli et al. 2004; Hein et al. 2007). STS preferably integrates familiar material, independent of congruency (Beauchamp et al. 2004; Hein et al. 2007). AV integration regions in IFC were preferably involved in the integration of familiar but incongruent material (e.g., dog picture and meowing sound; Taylor et al. 2006; Hein et al. 2007) or unfamiliar stimuli (Hein et al. 2007). Taken together, these findings indicate that the cortical activation patterns related to AV integration are affected by both the familiarity and semantic congruency of the stimuli in both modalities.

In everyday life there are many cases where the familiarity of an object's image and its sound changes, for example when we learn to associate a certain blip or error signal to a new technical device. Based on previous findings showing an impact of familiarity on neural AV integration it is plausible to assume that such training-induced familiarity modifies cortical AV integration. However, the neural plasticity of AV integration has not been investigated to a larger extent. In the present study, we used functional magnetic resonance imaging (fMRI) to investigate modulation of cortical AV integration after training of cross-modal associations between unfamiliar 3-dimensional object images and sounds.

A few previous fMRI studies on cross-modal learning have assessed time-dependent activation changes during AV learning (Gonzalo et al. 2000; Tanabe et al. 2005) or changes in functional connectivity between unimodal cortices (von Kriegstein and Giraud 2006). Cortical regions exhibiting plasticity correlating with AV learning have been found in the posterior hippocampus, parietal, and dorsolateral prefrontal regions (Gonzalo et al. 2000) and the STS (Tanabe et al. 2005). These latter studies employed low-level AV stimulus sets including musical chords or white noise for auditory and Chinese characters or 2-dimensional abstract patters for visual stimulation. In contrast to the 2-dimensional AV stimuli used by Gonzalo et al. and Tanabe et al., the unfamiliar object images used in our study had a 3-dimensional shape together with shadings, etc., resembling everyday life cases (e.g., images of unknown technical devices). Moreover, 3-dimensionality of our unfamiliar objects makes them also more comparable to familiar AV pairings, which in most cases are also comprised of a 3-dimensional object paired with a complex sound. So far, training-induced changes in AV integration for such object-like, *initially nonmeaningful stimuli* have not been investigated yet.

There is interesting evidence for training-induced neural changes in object-related sound processing (Amedi et al. 2007) as well as processing of 3-dimensional, initially unfamiliar, object images (James and Gauthier 2003; Weisberg et al. 2007). Amedi et al. compared neural activations in blind and sighted subjects after they had learnt to recognize everyday objects

© 2008 The Authors
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

based on complex artificial sound patterns. Training-induced changes of activation elicited by these sounds were found in intraparietal sulcus and prefrontal cortex, mainly along the precentral sulcus. After training, blind subjects revealed sound-induced activations in the lateral-occipital tactile–visual area, providing striking evidence for a cross-modal representation of 3D shape in this region (Amedi et al. 2007).

In one previous visual study subjects learnt linguistic labels ("names") for pictures of complex artificial stimuli (so-called "greebles"), with these labels referring either to movements or sounds (James and Gauthier 2003). After learning, the purely visual presentation of these artificial stimuli led to activations of the respective motion- and sound-processing cortical regions. In another study subjects were presented with artificial "tool-like" objects and trained on how to use these novel tools (Weisberg et al. 2007). Before training, the presentation of photographs of these objects only led to ventral visual activations, whereas the neural responses after training resembled those previously found for photographs of highly familiar common tools.

It is an open question whether there is similar neural plasticity in AV integration of cross-modally trained objects. To investigate this question, our participants were presented with artificial 3D object images (so-called "fribbles") and sounds (Fig. 1*A*) in a pretraining fMRI experiment (PRE), a cross-modal training session, and a subsequent second fMRI experiment (POST). In the training session, subjects learnt novel associations between the sounds and images, thus increasing their familiarity with these artificial stimuli. In a control condition, we presented images and sounds of highly familiar animals. The results of the PRE-training session have been reported elsewhere (Hein et al. 2007).

Our first prediction was that the cortical integration pattern for artificial AV pairings would change with training. We were recently able to demonstrate that familiar AV pairings are integrated in temporal regions such as the STG and STS, whereas integration of unfamiliar artificial object sounds and images mainly activated right IFC (Hein et al. 2007). Based on these findings, we expected that a training-induced increase of familiarity should correlate with an increase of integration-related AV activation in frontal and additional temporal regions. Our second prediction was that the establishment of associations between artificial sounds and images should also be reflected by effects of congruency on cortical activation patterns. We expected activation differences between

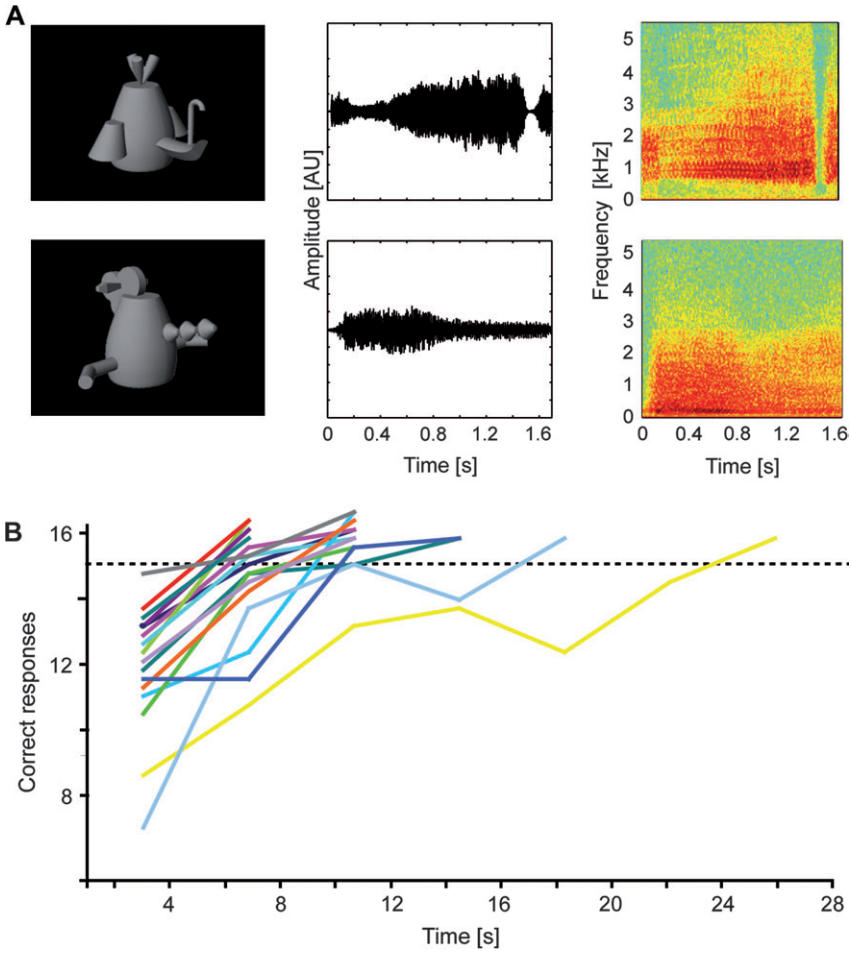

Figure 1. AV stimuli and behavioral data. (*A*) AV stimuli consisted of combinations of gray-scale photographs (left column) and complex sounds of both natural (i.e., animals) and artificial material (i.e., so-called "fribbles"). AV combinations could be either congruent, that is, corresponding to the learnt associations between the artificial visual and auditory stimuli (upper row in *A*) or incongruent (lower row in *A*). Middle and right columns provide the amplitude waveforms and spectrograms for typical exemplars of artificial sounds. (*B*) Behavioral training data of 17 subjects and interindividual differences in AV learning speed. The dashed horizontal line indicates the criterion of 15 (out of a maximum of 16) correct responses which had to be reached in 4 successive training blocks in order to terminate the training session. AU, arbitrary units.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

conditions with artificial AV pairings for which an association was acquired and those comprised of randomly paired artificial sounds and images. There are 2 different sets of regions, which have been found to show an AV activation profile sensitive to the semantic congruency between object images and sounds. One set of regions comprises higher-order auditory or visual regions, which are known to be involved in the integration of congruent sounds and images (Belardinelli et al. 2004; Hein et al. 2007). Accordingly, training-induced congruency effects should be found in such temporal regions. The other region is the IFC, which was found to integrate unfamiliar artificial AV pairings (Hein et al. 2007), and, moreover, showed stronger AV activations for incongruent as compared with congruent familiar AV combinations (Taylor et al. 2006; Hein et al. 2007). Based on the latter, we also expected IFC activation in the POST-training session in response to artificial AV combinations which, after training, were considered as mismatching.

Materials and Methods

Overall Structure of the Study

The study consisted of 3 major parts. All subjects underwent a first fMRI session (hereafter referred to as PRE), which served as a baseline for the remaining experiment. Subsequently the same subjects participated in a behavioral training session, which aimed at establishing an association between artificial auditory and visual stimuli. On the day following the training session, the second fMRI scan (POST) was performed with the same experimental setup as in the PRE session forming the third and final part of the study. Prior to the POST scan, subjects underwent behavioral testing to ensure that the learning criterion (15 out of 16 correctly identified AV combinations; see below) was still fulfilled. The delay between the PRE and the POST sessions varied between 74 and 123 days (mean: 94.9 days).

Participants

Eighteen adults (7 females, mean age = 29.8 years, range 23–41 years, one left-handed), with normal or corrected-to-normal vision and hearing participated in the study. They received information on MRI and a questionnaire to check for potential health risks and contraindications. Subjects gave their written informed consent in accordance with the declaration of Helsinki.

Stimuli

During the fMRI sessions, gray-scale 3D images were presented in the center of a black screen with a mean stimulus size of 14.6° visual angle. A fixation cross was constantly present. Artificial images were chosen from the "fribble" database (Fig. 1*A*; http://titan.cog.brown.edu:8080/TarrLab/stimuli). Sounds were presented via headphones simultaneously to both ears. They consisted either of animal vocalizations or artificial sounds, created through distortion (played backwards and filtered with an underwater effect) of the same animal vocalizations. In a behavioral pretest, distorted sounds were presented to 8 subjects. None of the distorted sounds was associated with any common object. Sounds and images were presented in stimulation blocks at a rate of one per 2 s. Stimulation blocks consisted of 8 stimulus events.

Procedure for the fMRI Experiments (PRE and POST)

The fMRI sessions were comprised of 8 different experimental conditions: unimodal conditions for animal vocalizations (A-NAT), artificial (i.e., distorted) sounds (A-ART), animal images (V-NAT), and artificial fribble images (V-ART). Bimodal conditions consisted of synchronously presented sounds and images with varying degrees of semantic congruency (AV-NAT-CON: e.g., dog photograph—barking sound; AV-NAT-INCON: e.g., dog photograph—meowing sound; AV-ART-CON, and AV-ART-INCON). Note that in the first fMRI session (PRE), that is, prior to the establishment of learnt associations between the artificial sounds and images, no differences between AV-ART-CON and AV-ART-INCON were expected. For the PRE session, the sole difference between these 2 experimental conditions was that the AV stimuli were presented either in fixed or in randomized pairings.

The 8 conditions were presented within a block design with 16 s of stimulation (8 recording volumes) alternating with 16 s of central fixation. A complete experimental run consisted of 2 cycles of all experimental conditions with additional 8 volumes of fixation at the beginning of the run. We had 5 experimental runs, with the order of blocks being pseudorandomized in each run and counterbalanced across runs. In the fMRI sessions, subjects were instructed to fixate and attentively perceive the experimental stimuli.

A passive paradigm was chosen to minimize task-related activations of the frontal cortex (see also Calvert et al. 2000; Belardinelli et al. 2004; van Atteveldt et al. 2004, 2007). Given the sluggishness of the blood oxygenation level–dependent (BOLD) fMRI signal, it might otherwise be hard to disentangle a potential involvement of frontal regions in AV integration from the task-related frontal activations.

AV Association Training

Prior to the experiments, 8 AV associations—each consisting of one gray-scale fribble image and one distorted sound—were declared as "correct." Training consisted of several runs, each containing 4 blocks with 16 trials that were presented in randomized order to reduce predictability. Eight of these trials comprised the "correct" associations; in the other 8 trials the artificial visual and auditory stimuli were paired randomly. Subjects had to categorize the AV pairings as "correct" or "incorrect" via button press after each trial. Because they performed at a self-paced speed, only accuracy data were analyzed. Immediately after pressing the button, a visual feedback was given on the screen ("The answer was correct." or "The answer was incorrect."). After each block the number of correct trials was displayed as an additional feedback. Subjects were instructed to avoid learning strategies based on verbalization, for example by using invented "names" for the stimuli. The number of runs was not fixed because the aim was to ensure that each subject received the necessary amount of training in order to successfully learn the "correct" associations. The criterion was reached when in 4 successive blocks within one run at least 93.75% (i.e., 15/16) of the answers were correct. If a subject did not meet this criterion, another run of 4 blocks started. At the beginning of this new set of blocks as well as at the very beginning of the whole behavioral training session the "correct" associations were presented to further facilitate the training process.

MRI Data Acquisition

fMRI was performed with a 3 Tesla Magnetom Allegra scanner (Siemens, Erlangen, Germany). 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.

Whole-Brain Data Analysis

Data were analyzed using the BrainVoyager QX software package (version 1.8; Brain Innovation, Maastricht, The Netherlands). The first 4 volumes of each experimental run were discarded to preclude T_1 saturation effects. Preprocessing of the functional data included the following steps: 1) 3-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 kernel of 8-mm full-width at half-maximum.

Volume-based statistical analyses were performed using random-effects general linear models (RFX GLM; 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. RFX GLMs were computed both separately for each

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

experimental session and across sessions. The latter model permitted explicit testing for effects across time.

Neural correlates of object-related AV integration were assessed for artificial material in the PRE and POST fMRI sessions. We searched for regions that were 1) significantly activated during each of the unimodal conditions (A; V), and 2) responded more strongly to bimodal AV stimulation than to each of its unimodal components. Accordingly, the identification of brain regions involved in object-related AV integration was based on significant activation in a $0 < A < AV > V > 0$ conjunction analysis, a statistical criterion well-established in previous studies (Beauchamp et al. 2004; van Atteveldt et al. 2004, 2007; Hein et al. 2007).

In addition, we directly tested for PRE–POST effects on the basis of the whole-brain RFX GLM across both data sets. The respective conjunction analysis required both significantly stronger responses to bimodal artificial AV conditions compared with both unimodal conditions (A-ART and V-ART) as well as stronger responses to bimodal artificial conditions during the POST versus the PRE session (AV-ART-POST > PRE).

In a second step of analysis we tested for effects of congruent and incongruent AV stimulation. We reasoned that training might have induced such effects for artificial AV material in cortical regions which are also sensitive to congruency relations in natural AV stimuli. Accordingly, we searched for voxels exhibiting stronger signals for the contrast AV-ART-CON > INCON (or vice versa) in those regions which showed additionally significant results in the contrast AV-NAT-CON > INCON (or vice versa). Statistically this corresponds to a main effect of *congruency* for both types of stimuli. To explicitly test this effect, we entered the beta values of the 4 AV conditions from the RFX GLM of the POST fMRI session into a (whole-brain) 2-factorial ANOVA as implemented in the respective Brainvoyager QX tool. Factors (and levels) were *stimulus type* (NAT, ART) and *congruency* (CON, INCON). The main effect of *congruency* was then specified in this model by contrasting CON versus INCON pooling across the factor *stimulus type*. Voxels identified by these contrasts were taken as regions of interest (ROIs) and further analyzed with respect to the question which response patterns they exhibited in the first experimental session (see below). This aimed at further supporting potential training-induced effects for artificial AV stimulation in the POST session.

In addition to these contrasts, we also searched in unimodally defined sensory cortices for training-induced effects of both AV integration and congruency. To this end, we contrasted responses to each of the natural unimodal conditions (A-NAT; V-NAT) separately with responses to the respective artificial counterpart (A-ART; V-ART). We chose these contrasts to identify those unimodal regions which are likely to be sensitive to semantic features of object stimuli, or, more precisely, regions corresponding to higher-order auditory cortices and the ventral visual cortex. Voxels determined by these contrasts were similar to the previously described regions further analyzed in a ROI analysis.

Generally, 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 $t > 4$ ($P < 0.001$, uncorrected; unless otherwise indicated) and was then submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates. After 1000 iterations the minimum cluster-size that yielded a cluster-level false-positive rate of 5% was used to threshold the statistical map.

When possible, group-averaged functional data were projected on inflated representations of the left and right cerebral hemispheres of one subject. As 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.

ROI-based Analysis

Regions identified in the previous whole-brain analyses were further investigated in the course of a ROI-based analysis with respect to the question of congruency effects induced by novel AV associations. In addition to the voxels identified by the main effect of *congruency* in the data of the POST fMRI session we defined further ROIs based on the respective unimodal (i.e., auditory and visual) contrasts between natural and artificial conditions. We conducted RFX GLMs and computed *t*-tests in these ROIs comparing responses to congruent and incongruent AV stimulation for the natural and the artificial conditions in both sessions. We were mainly interested in regions exhibiting a stable effect for the comparison of congruent and incongruent natural stimuli in both experimental sessions and, additionally, a novel, that is, training-induced effect of congruency for artificial AV stimuli only in the POST session. Statistically, this would be most robustly reflected by a transition from a significant interaction of the factors *material* and *congruency* in the PRE session (due to congruency or incongruency effects only for the natural material) to a main effect of *congruency* in the POST session. Additionally, 2-factorial ANOVAs with the factors *time* (PRE, POST) and *congruency* (CON, INCON) were computed separately for the natural and the artificial AV stimuli to verify that changes occurred only for the latter. Furthermore, we entered the ROI beta weights of the GLM including data from both experimental sessions into a 3-factorial ANOVA with the factors *material* (NAT, ART), *congruency* (CON, INCON), and *time* (PRE, POST). We thereby explicitly tested for any significant interactions of these 3 factors which should most likely reflect changes in the congruency relation of artificial stimuli across time.

In all figures containing bar plots of ROI-based beta estimates mean standard errors are additionally reported. The latter were corrected for intersubject effects similar to previous studies (Altmann et al. 2007; Doehrmann et al. 2008). In particular, the fMRI responses were calculated individually for each subject by subtracting the mean percent signal change for all conditions within this subject from the mean percent signal change for each condition and adding the mean percent signal change for all the conditions across subjects.

Evaluation of Repetition-Related Effects

In a final analysis step we aimed at determining to which extent effects for artificial AV stimuli might be affected by unspecific repetition-related effects, for instance due to adaptation. To this end, we computed in the across-session GLM the contrast AV-ART-CON (POST) > AV-ART-INCON (PRE) and vice versa. The rationale for this was that the stimuli of the experimental condition AV-ART-CON were repeated most frequently because they were presented in this constellation in both fMRI sessions as well as in the training session. Thus, any repetition-related effects were most likely present in fMRI signals of this experimental condition. On the other hand, the condition AV-ART-INCON (PRE) is least likely to contain any repetition-related effects due to 1) the randomization of the relation between artificial stimuli and 2) the fact that the data were recorded in the first experimental session.

To further corroborate the robustness of our other results, we increased the likelihood of finding any unspecific repetition-related effects by using uncorrected thresholds of $t(17) = 4$ and $t(17) = 3$ for these contrast maps. These maps were then projected to inflated representations of the cerebral cortex of one of our subjects. We then also projected the main effect of incongruency onto these cortical surfaces to evaluate the potential overlap of these effects with the former one. Furthermore, we additionally computed ROI-based paired *t*-tests for AV-ART-CON (POST) > AV-ART-INCON (PRE) to exclude unspecific repetition effects for the trained material in the regions exhibiting our most relevant experimental results. And finally, we also contrasted AV-NAT-CON (POST) > AV-NAT-CON (PRE) in order to control for stimulus-specific repetition effects regarding the untrained, but still repeated natural material.

Results

Behavioral Training Results

In the training period, subjects were presented with combinations of artificial object sounds and images until they reached

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

a criterion of 15 out of 16 correct responses. The mean number of training blocks necessary to reach this criterion was 12.7 (range: 8–28) which was approximately equivalent to 3 runs (consisting of 4 blocks each) per subject. After the first run the mean number of correct responses was 11.63 (of 16 possible). At the end of the training session (i.e., after reaching the criterion) the number of correct trials increased to 15.25 per block in the final run. The general trend of increasingly successful associations between artificial AV stimuli was well-established for all subjects (Fig. 1*B*). Because prior to each training run (consisting of 4 blocks each) the correct AV associations were presented to the subjects, the results for the first blocks were already well above chance level. The same held true for the elevated performance after each run, which was probably due to retrieval from short-term buffers, rather than an establishment of long-term associations between the stimuli. Our criterion for a stable performance of 15/16 correct responses on average over 4 successive blocks was chosen to account for this.

On the day of the POST fMRI session, we checked if subjects still met the criterion. For this purpose, the same procedure as in the training session was repeated. Most of the subjects (i.e., 15/18) could immediately replicate their performance from the day before. Only 3 of them needed another run of 4 training blocks to reach the criterion again. The mean number of blocks needed for refreshment was 4.89 with a mean performance of 15.41 correct responses per block. This shows that for each subject the trained artificial AV associations were still present at the time of the POST fMRI session.

fMRI Results

Whole-Brain Data Analysis

To investigate whether the training of crossmodal associations between formerly unfamiliar artificial object images and sounds had an impact on AV integration-related cortical activations, we analyzed the data of the PRE and POST fMRI sessions separately. For this purpose, we employed the so-called MAX-criterion (i.e., AV > max [A, V]).

PRE Training

The results of the PRE fMRI session have already been reported in detail in our previous publication (see Hein et al. 2007). As can be seen there, unfamiliar artificial images and sounds were mainly found to be integrated in right IFC (Fig. 2*A*), whereas familiar natural AV objects were integrated in a bilateral fronto-temporal network of regions including IFC, STS, and STG. Note

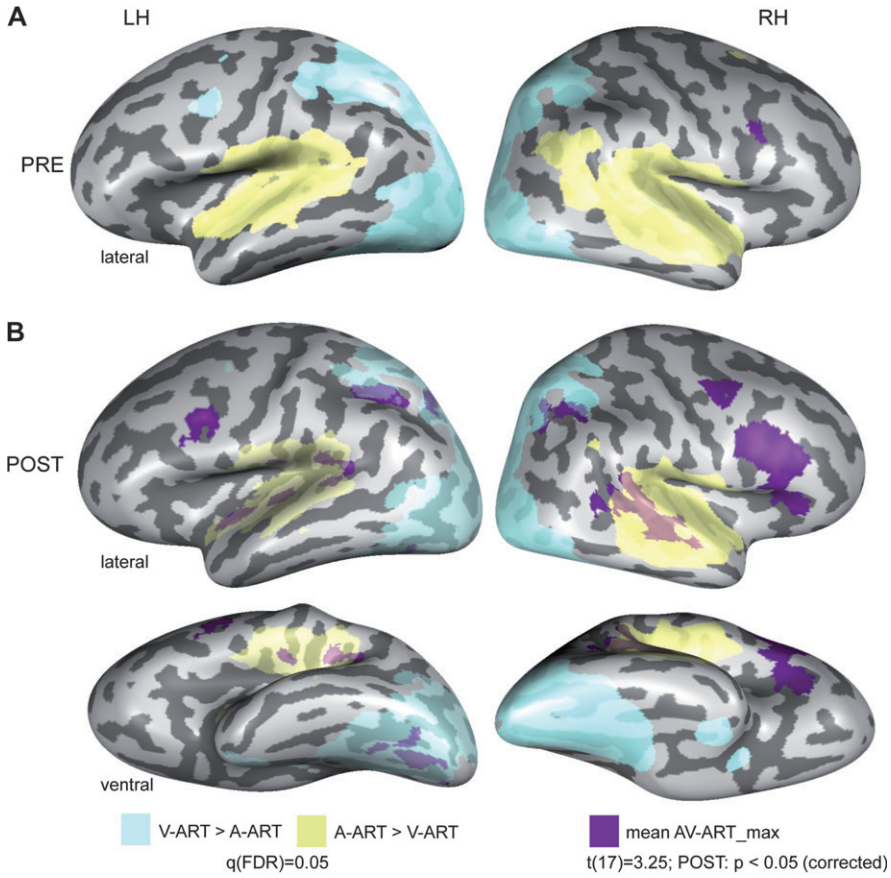

Figure 2. AV integration effects for artificial material. Cortical regions involved in AV integration of artificial stimuli during the first (*A*) and the second (*B*) experimental session were determined by a conjunction analysis comparing the mean BOLD fMRI responses to both (congruent and incongruent) artificial AV conditions to each unimodal condition (MAX-criterion: AV-ART > max[A-ART,V-ART]). Similar to our previous study (Hein et al. 2007) the additional criteria A-ART > 0 and V-ART > 0 were added to this conjunction. Maps were then adjusted to an initial *t*-value of 3.25 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at $P < 0.05$ (corrected minimum; cluster-size 189 voxels). These corrected maps were then projected onto inflated reconstructions of both cortical hemispheres of one subject from our sample. See Table 1A for Talairach coordinates and extents of cortical clusters depicted here. Furthermore, unimodal activations as determined by the contrasts A-ART > V-ART (auditory; yellow) and V-ART > A-ART (visual; light blue) were computed and corrected based on the false-discovery rate (FDR) yielding a significance value of $q(\text{FDR}) < 0.05$. LH and RH denote the left and right hemispheres, respectively. PRE and POST refer to the first and second experimental session.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

Table 1
Talairach coordinates of regions showing significant effects

Cortical region	x	y	z	Cluster size (number of voxels)
A. Max-contrast				
Left hemisphere				
STG (posterior)	−46	−36	13	799
STG (middle)	−44	−22	7	329
STG (anterior)	−41	−10	−4	344
Ventral occipito-temporal cortex (combined)	−35	−60	−17	755
Inferior frontal gyrus	−46	5	26	464
Inferior parietal lobule	−31	−50	37	815
Precuneus	−23	−66	35	341
Posterior cingulate	0	−28	23	487
Right hemisphere				
Posterior STG	52	−31	9	2742
Middle frontal gyrus (ventral)	44	14	21	2696
Middle frontal gyrus (dorsal)	42	−2	48	523
Inferior parietal lobule (anterior)	31	−52	39	372
Precuneus	28	−62	34	198
Posterior cingulate	1	−28	23	556
B. PRE–POST comparison				
Left hemisphere				
Middle frontal gyrus (ventral)	−41	21	22	522
Inferior frontal gyrus	−45	5	26	465
Insula	−29	21	4	205
Inferior parietal lobule	−31	−50	38	399
Posterior cingulate	0	−29	21	374
Right hemisphere				
Inferior frontal gyrus	47	15	21	1344
Precentral gyrus	49	4	31	319
Insula	33	21	3	554
Inferior parietal lobule	34	−52	38	355
Presupplementary motor area	7	9	50	222
Posterior cingulate	1	−28	22	410

Note: (A) For the max-contrast (POST) and (B) for the PRE–POST comparison. Clusters were determined by comparing the mean of the 2 artificial AV conditions to the respective unimodal controls (volume-based cluster threshold of the map: 189 voxels; surface-based threshold 25 mm^2). Cluster sizes for the cortical regions reported in Figure 2*B* (part A of the table) and in Figure 3 (part B) are provided in a resolution of $1 \times 1 \times 1$ mm^3 voxels.

that the results of the PRE session are shown with a slightly higher initial statistical threshold ($t(17) = 3.25$) than in the previous report of the data (Hein et al. 2007; $t(17) = 3$) before correction for multiple comparisons. As expected, the respective unimodal contrasts A-ART > V-ART (yellow) and the reverse contrast (light blue) revealed robust activations in visual and auditory cortex.

POST Training

We tested for potential plasticity of cortical AV integration as defined by the max-criterion (i.e., AV > max [A, V]) for artificial AV pairings (pooled across congruent and incongruent conditions) in the POST fMRI session. The main results of this analysis ($t(17) = 3.25$, $P < 0.05$, corrected, cluster-size threshold = 189 mm^3) are depicted in Figure 2*B* (dark blue-colored patches), whereas the respective Talairach coordinates and cluster sizes of these regions are provided in Table 1A. The AV integration-related cortical network included bilateral IFC and STG/STS regions, in line with our first prediction. Moreover, AV integration-related activity was found in bilateral anterior insula and IPL, as well as precuneus and VOT in the left hemisphere.

In addition, we directly tested for PRE–POST effects on the basis of a whole-brain RFX GLM across both fMRI data sets. The respective conjunction analysis included the following 3 criteria: 1) AV-ART-POST > A-ART-POST, 2) AV-ART-POST > V-ART-POST, and 3) AV-ART-POST >AV-ART-PRE. The main results of this analysis ($t(17) = 3.25$, $P < 0.05$, corrected, cluster-size threshold = 189 mm^3) are depicted in Figure 3, whereas the respective Talairach coordinates and cluster sizes of these regions are provided in Table 1B. The results revealed activations in bilateral posterior cingulate regions and fronto-parietal

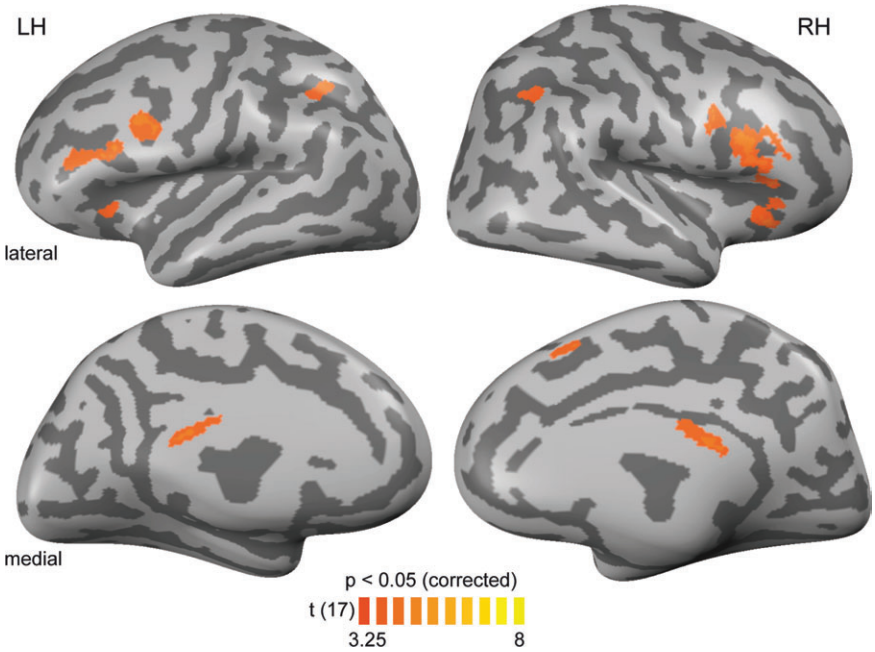

Figure 3. PRE–POST comparison. Results of a contrast investigating effects from the first (PRE) to the second (POST) fMRI session on the basis of beta weights from a whole-brain RFX GLM across both data sets. A conjunction analysis required both significantly stronger responses to bimodal artificial AV conditions compared with both unimodal conditions (A-ART and V-ART) as well as stronger responses to bimodal artificial conditions during the POST versus the PRE session (AV-ART-POST > PRE). Maps were adjusted to an initial *t*-value of 3.25 and a volume-based cluster-size threshold of 189 voxels (see Fig. 2), resulting in a new map thresholded at $P < 0.05$ (corrected). These corrected maps were then projected onto inflated reconstructions of both cortical hemispheres of the same subject as in Figure 2. See Table 1B for Talairach coordinates and extents of cortical clusters depicted here.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

(IPL, IFC), but not temporal portions of the aforementioned AV integration-related network.

In our second prediction we hypothesized that the training of crossmodal associations between formerly unrelated sounds and images might also induce effects of congruency, that is, activation differences in 2 possible directions: AV-ART-CON > INCON or AV-ART-INCON > CON. To test whether our training procedure had established congruency effects for our artificial AV stimuli corresponding to those for familiar natural stimuli, we analyzed the POST session data by means of a whole-brain ANOVA with the factors *material* (NAT, ART) and *congruency* (CON, INCON). This analysis revealed a statistical main effect ($t(17) = 4$, $P < 0.05$, corrected) for incongruency in left anterior IFC (Fig. 4*A*; see Table 2 for Talairach coordinates). A further ROI-based analysis of the activation profiles (by means of BOLD-signal time courses and

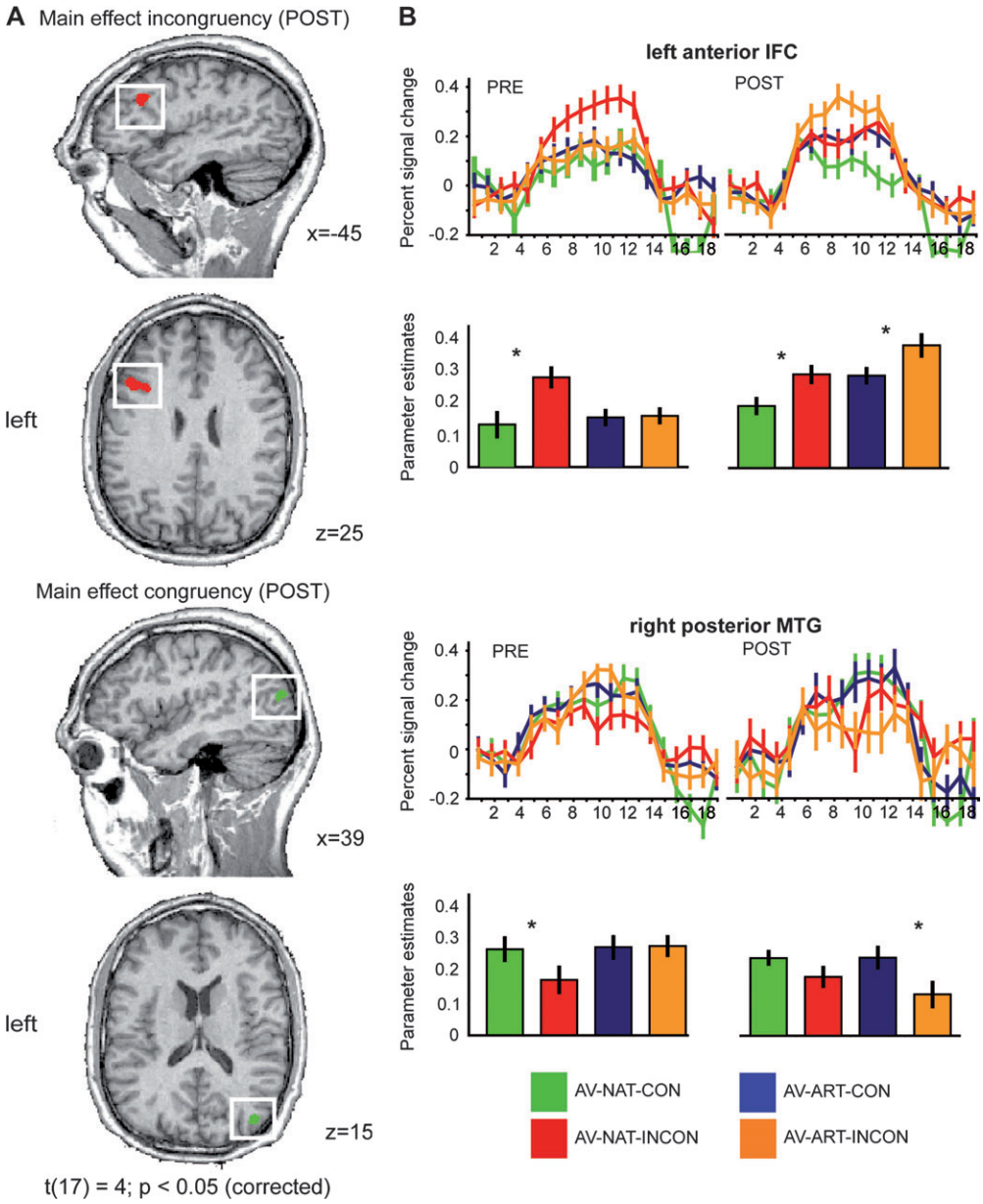

Figure 4. POST-training main effects of incongruency and congruency. Beta weights from the whole-brain RFX GLM of the POST-training session were entered into a 2-factorial ANOVA with the factors *stimulus type* (NAT, ART) and *congruency* (CON, INCON). In this model we specified a contrast corresponding to the main effect of *congruency* because the latter optimally captured the assumed development of congruency and incongruency for artificial AV material in those cortical regions which were similarly sensitive to *congruency* in natural AV stimuli. (*A*) Results of the contrasts are shown in the left panel in the direction of incongruency (INCON > CON) and congruency (CON > INCON). Maps were adjusted to an initial *t*-value of 4 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at $P < 0.05$ (corrected; cluster-size 115 voxels). Maps were then projected on sagittal and transversal slices of an anatomical dataset of the same subject which provided the cortex reconstructions in Figure 2. (*B*) ROI-based analyses of cortical regions depicted in (*A*). Significant clusters in the left anterior IFC (upper panel) and right posterior middle temporal gyrus (pMTG; lower panel) were further analyzed with regard to the question of congruency or incongruency effects. For each ROI the respective contrasts (e.g., AV-NAT-INCON > AV-NAT-CON) were calculated in separate RFX GLMs for both types of stimulus material and for both sessions. Time course information and beta estimate plots are provided for these 2 regions. Error bars represent mean standard errors. Asterisks denote significant effects on a level of $P < 0.05$. Color coding is provided at the bottom of the figure. Additional statistical information can be found in Table 2.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

Table 2
Summary of ROI analysis on congruency and incongruency effects

Cortical region	AV-NAT-INCON > CON	AV-ART-INCON > CON	Material × congruency	*Material × Time × Congruency*
Left IFC (−38, 13, 29)				
POST	$t(17) = 3.063$; $\boldsymbol{P = 0.007042}$	$t(17) = 3.244$; $\boldsymbol{P = 0.004771}$	$t(17) = 0.098$; $P = 0.923029$	$F_{1,17} = 11.310$; $\boldsymbol{P = 0.004}$
PRE	$t(17) = 3.021$; $\boldsymbol{P = 0.007710}$	$t(17) = 0.235$; $P = 0.817001$	$t(17) = 2.929$; $\boldsymbol{P = 0.009363}$	
	AV-NAT-CON > INCON	*AV-ART-CON > INCON*	*Material × Congruency*	*Material × Time × Congruency*
Right pMTG (40, −78, 18)				
POST	$t(17) = 2.027$; $P = 0.058627$	$t(17) = 4.201$; $\boldsymbol{P = 0.000601}$	$t(17) = 1.851$; $P = 0.081568$	$F_{1,17} = 4.086$; $P = 0.059$
PRE	$t(17) = 2.328$; $\boldsymbol{P = 0.032499}$	$t(17) = 0.178$; $P = 0.860894$	$t(17) = 1.718$; $P = 0.103925$	
Right FG (35, −40, −16)				
POST	$t(17) = -0.322$; $P = 0.751495$	$t(17) = 2.340$; $\boldsymbol{P = 0.031751}$	$t(17) = -1.622$; $P = 0.123256$	$F_{1,17} = 0.101$; $P = 0.754$
PRE	$t(17) = -0.793$; $P = 0.438727$	$t(17) = 0.628$; $P = 0.538534$	$t(17) = -0.849$; $P = 0.407857$	

Note: Clusters were determined by computing main effects of AV congruency and incongruency in the 2-factorial ANOVA on the factors *Material* (NAT, ART) and *Congruency* (CON, INCON). Volume-based cluster threshold of the map was 115 voxels; surface-based threshold was 25 mm^2. Coordinates refer to Talairach space (Talairach and Tournoux 1988). pMTG: posterior middle temporal gyrus.

beta estimates; Fig. 4*B*) demonstrated that incongruency effects for familiar natural stimulus material were stable across the PRE and POST fMRI sessions. In contrast, similar incongruency effects for the artificial AV objects were only observable during the POST fMRI session, that is, after successful crossmodal association training. This was also evident in a significant *time* × *congruency* interaction ($F_{1,17} = 5.042$, $P < 0.05$) for the artificial material only. Additionally, these training-induced shifts in left IFC were also indicated by the transition from a significant statistical interaction of the factors *material* and *congruency* in the PRE session (driven by the response to incongruency in familiar natural AV stimuli) to a main effect of incongruency which was constitutive for this ROI. A summary of these analyses is given in Table 2.

The reverse contrast to this examination of training-induced incongruency, that is, a main effect of congruency, revealed a cluster in right posterior middle temporal gyrus (MTG). The same ROI analysis as for the left IFC demonstrated that a stronger response to congruent as compared with incongruent artificial AV conditions was only present in the POST fMRI session. This was again supported by a significant *time* × *congruency* interaction ($F_{1,17} = 9.764$, $P < 0.05$). However, a corresponding congruency effect for familiar natural stimuli was only significant in the PRE fMRI session and just failed to reach significance in the POST session ($t(17) = 2.027$, $P = 0.059$). Therefore, also the *material* × *congruency* interaction was not significant in this cluster ($t(17) = 1.851$, $P = 0.082$; see Table 2 for a complete summary).

Additionally, we conducted a 3-factorial ANOVA across the data of both sessions with the factors *material* (NAT, ART), *time* (PRE, POST), and *congruency* (CON, INCON). This analysis revealed a significant *material* × *time* × *congruency* interaction in left aIFC which was largely driven by the change in the fMRI response to the AV-ART-INCON condition. In line with our above reported analysis the corresponding effect in pMTG was less pronounced as revealed by a trend toward significance of this interaction (see Table 2 for the respective *F* and *P* values).

In line with our prediction this analysis showed that left anterior IFC was more strongly activated in response to formerly unrelated artificial sounds and images. After training these IFC responses were more pronounced for stimulation with incongruent than congruent pairs of artificial AV stimuli, whereas right posterior MTG showed a complementary, but less pronounced congruency preference pattern. These effects resemble the congruency effects for familiar natural AV objects which have been found consistently across both sessions in these regions.

Visual and Auditory ROIs

Based on previous results (Belardinelli et al. 2004; van Atteveldt et al. 2004; Hein et al. 2007), we also predicted congruency effects in higher–level visual and auditory regions. To define higher-level visual and auditory regions involved in the semantic processing of sounds and images, we contrasted the respective natural and artificial conditions (i.e., A-NAT > A-ART and V-NAT > V-ART, respectively; see Materials and Methods). This analysis revealed significant activations in bilateral STG and right fusiform gyrus (FG), which were then used as ROIs (Fig. 5). In the right FG (Fig. 5*A*), the crossmodal association training induced a significant congruency effect, that is, stronger AV activation to congruent as compared with incongruent artificial material (PRE: $t(17) = 0.793$, $P = 0.44$; POST: $t(17) = 2.340$, $P < 0.05$), without corresponding congruency effects for natural stimuli (see Table 2 for details). Table 2 again summarizes the outcome of these ROI analyses.

In contrast, unimodal auditory regions in STG showed no significant AV congruency effect for artificial material (Fig. 5*B*). However, right STG revealed a training-induced AV integration effect for artificial material. The latter was characterized by a significantly stronger response to trained artificial AV stimuli than to each of the unimodal (i.e., auditory and visual) conditions in the POST session (PRE: $t(17) = 1.400$, $P = 0.17$; POST: $t(17) = 2.551$, $P < 0.05$).

Evaluation of Repetition-Related Effects

One alternative explanation for the stronger POST-training activity in IFC, IPL, and temporal regions could be based on the potential presence of unspecific repetition effects. This would reflect the fact that subjects have been stimulated with the artificial material in the PRE training, the behavioral training, and the POST-training sessions. In order to control for these potential repetition-related effects we contrasted the experimental condition which was arguably most likely to be affected by such effects (AV-ART-CON [POST]) with the experimental condition least likely to be affected in this regard (AV-ART-INCON [PRE]; see Materials and Methods).

Significant effects ($t(17) = 4$; $P < 0.001$, uncorrected) for the contrast AV-ART-CON [POST] > AV-ART-INCON [PRE] are shown in Figure 6*A* (orange colors). Particularly regions of the medial cortex located in the bilateral posterior cingulate and presupplementary motor area and the right precuneus were detected by this contrast. Only small clusters were revealed in the inferior-most and anterior frontal cortex. Most importantly, none of these clusters was overlapping with our main effect of incongruency reported in the previous analysis (white circle in

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

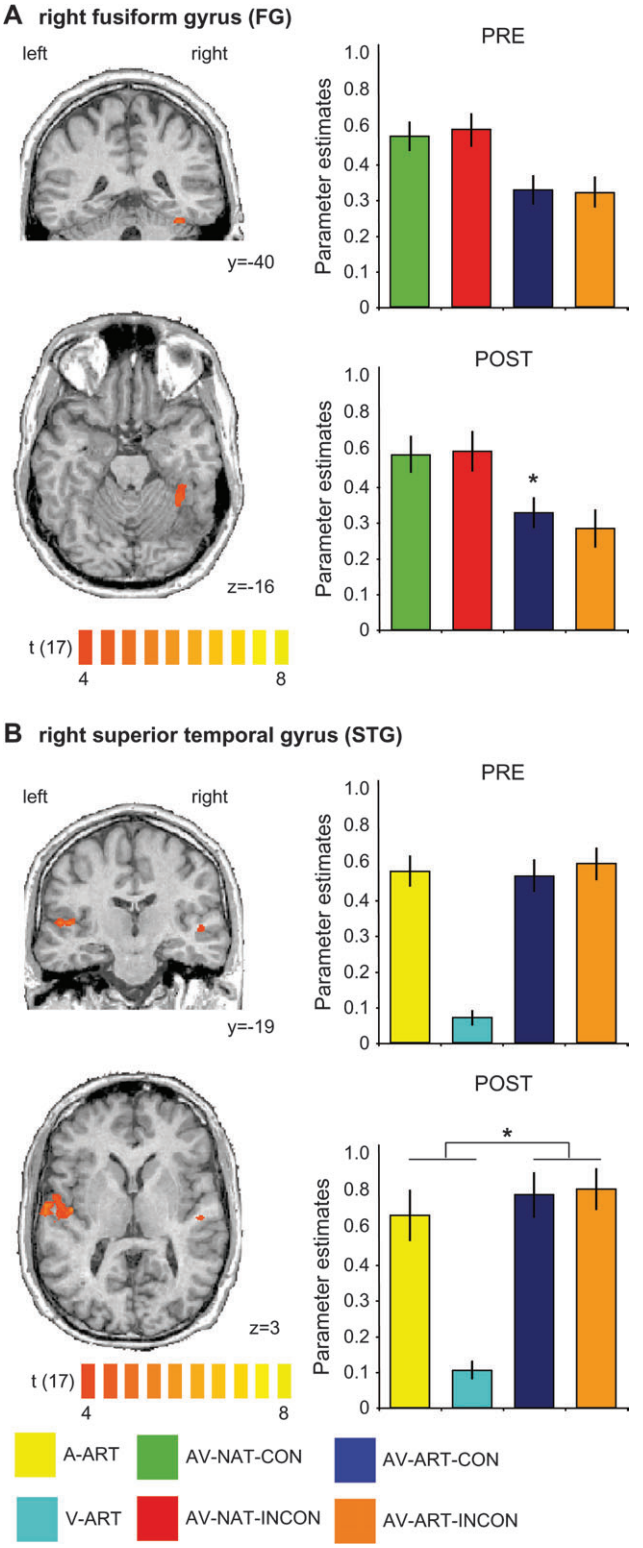

Figure 5. Unimodally defined ROIs: training-induced congruency and multisensory integration effects. Significant voxels are depicted for unimodal visual (*A*) and auditory (*B*) contrasts directly comparing stimulations with natural objects to those with artificial material in the POST fMRI session (i.e., V-NAT > V-ART (POST) and A-NAT > A-ART (POST), respectively). Maps were adjusted to an initial *t*-value of 4 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at $P < 0.05$ (corrected) for the visual (minimum cluster threshold: 271 voxels) and the auditory (minimum cluster threshold: 136 voxels) contrasts, respectively. The panels on the left show coronal (upper rows in *A* and *B*) and transversal slices (lower rows in *A* and *B*) of the same anatomical dataset used in Figure 4. The same ROI-based analysis as described in Figure 4 was applied to the right FG (*A*) for all bimodal conditions and for both fMRI sessions. A similar ROI-based analysis was applied to the region in right STG (*B*), revealing a training-induced AV integration effect (0 < A < AV > V > 0) for artificial material. Error bars of the beta estimate plots represent mean standard errors. Asterisks denote significant effects on a level of $P < 0.05$. Color coding is provided at the bottom of the figure. Additional statistical information for the FG ROI can be found in Table 2.

Fig. 6*A*). We chose an uncorrected threshold to maximize the likelihood of detecting such overlaps. However, even when further lowering the threshold for this map to $t(17) = 3$ no overlap with either the IFC (incongruency) or the right pMTG (congruency) could be detected (data not shown). Even ROI-based testing for repetition effects in these regions did not reveal any significant effects (for details see Table 3).

Finally, when contrasting AV-NAT-CON-POST versus PRE (Fig. 6*B*) in order to control for stimulus-specific repetition effects, significant activations ($t(17) = 3.25$; $P < 0.005$, uncorrected) were only found in a left lateral-occipital region (POST > PRE; orange-colored region in Fig. 6*B*) and bilateral medial occipital regions (POST < PRE; blue-colored regions). As none of these regions overlapped with any of our main ROIs, these results render it unlikely that the reported training-induced congruency and integration effects were due to stimulus repetition.

Discussion

In the present study we investigated training-induced cortical plasticity of object-related AV integration. We compared the neural correlates of AV integration for unfamiliar artificial material before and after a behavioral training session, in which subjects acquired crossmodal associations between object images and sounds. We assumed that our training session might modify neural AV processing in 2 different ways. First, we expected that the training-induced increase of familiarity might lead to an increase in integration-related AV activations. Second, we predicted that our training might also induce certain congruency effects as reflected in POST-training activation differences between trained artificial AV pairings and those which were paired randomly.

For the trained artificial AV stimuli, the POST-training results showed extended integration-related IFC activations bilaterally, and a recruitment of additional posteror parietal cortex (PPC) and temporal (right STG/STS, and left FG) cortical regions. Furthermore, training-induced effects of congruency were found for these artificial AV stimuli in several cortical regions. We revealed a training-induced AV incongruency effect in left aIFC and a complementary congruency effect in a posterior portion of right MTG. Both cortical regions showed similar effects for familiar natural AV stimuli in each of the 2 fMRI sessions.

Integration-Related AV Activations and Incongruency Effects

We found training-induced changes of AV integration-related activation patterns for (initially unfamiliar) artificial object stimuli, indicating cortical plasticity of object-related AV integration. Confirming our first prediction, crossmodal association training resulted in enhanced integration-related AV activation in a fronto-parieto-temporal network including cortical regions in bilateral IFC, IPL, STG/STS, and left FG. Within this network frontal regions showed the most robust

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

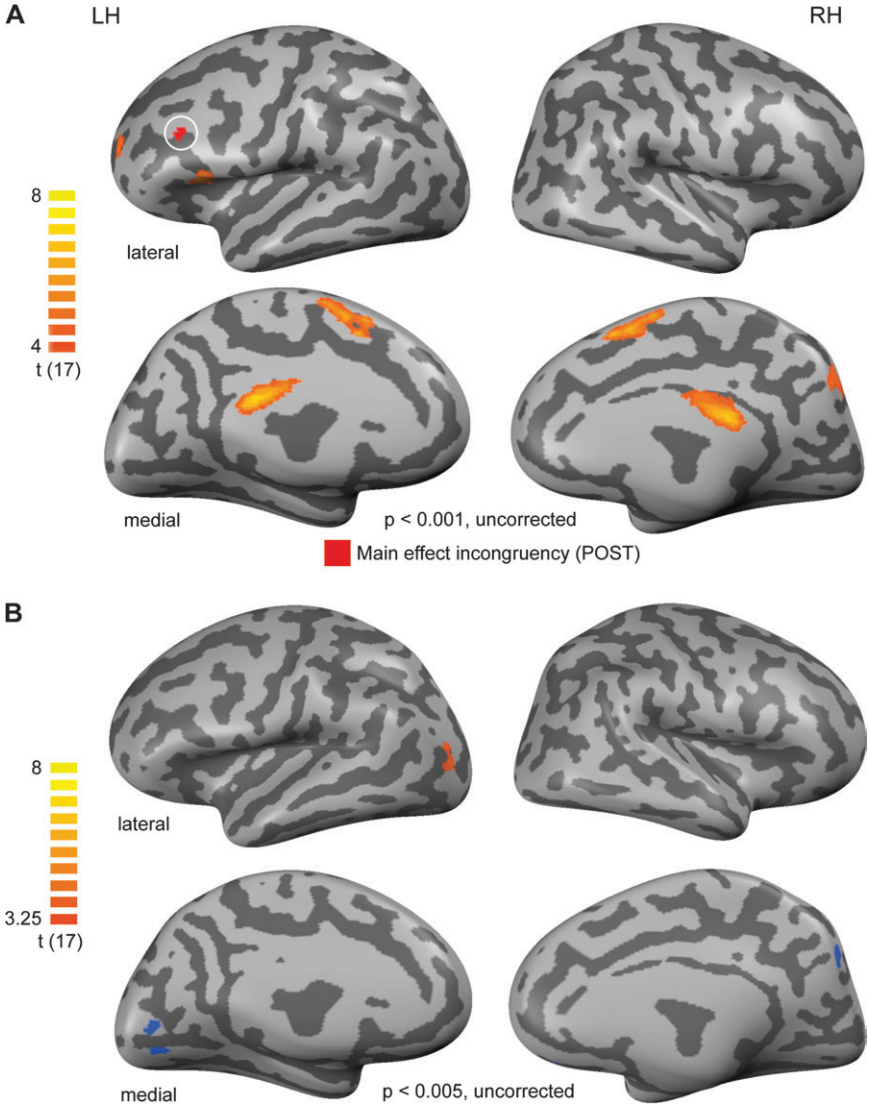

Figure 6. Repetition-related PRE–POST effects. (*A*) To control for repetition-related effects we contrasted the experimental condition which is arguably most likely to be affected by such effects (AV-ART-CON [POST]) with the experimental condition least likely to be affected by this (AV-ART-INCON [PRE]). Only positive effects (signal increases from PRE to POST) were significant at the chosen threshold. These are shown in an orange-to-yellow color code. Maps are thresholded to a *t*-value of 4 (uncorrected for multiple comparisons) to increase the sensitivity for potential repetition-related effects and were then projected onto the same cortex reconstructions used in Figure 2 shown in lateral (upper panels) and medial (lower panels) views. Only a minimal surface-based cluster threshold of 25 mm^2 was applied. Additionally, the main effect of *incongruency* as described in Figure 4 is shown (in red and highlighted by a white circle) to evaluate a potential frontal overlap of these maps. LH and RH denote the left and right hemispheres, respectively. (*B*) To control for stimulus-specific repetition effects we additionally contrasted the unlearned, but also repeated experimental condition comprised of natural AV stimuli. The corresponding contrast can be stated as follows: AV-NAT-CON (POST) > AV-NAT-CON (PRE). Because no effects were present on a statistical *t*-value of 4, we lowered the threshold to an uncorrected *t*-value of 3.25. The same anatomical basis as in the previous figures was used for the projection of results.

training-induced AV plasticity effects, reflected by significant POST > PRE effects. So far, training-induced changes in cortical processing of complex object stimuli had only been shown in the visual modality (James and Gauthier 2003; Weisberg et al. 2007).

Training-induced activation increases in brain regions which have been activated already before training, such as right IFC in our study, are a well-known result from studies investigating both AV association learning (Gonzalo et al. 2000) and training-induced effects in other domains, for example, perceptual processing (Schwartz et al. 2002), memory (Nyberg et al. 2003; Olesen et al. 2004) or learning of complex 3D object-related sounds (Amedi et al. 2007). Across studies investigating learning processes in diverse contexts, such effects were found in similar inferior frontal regions as those activated in our study, often in concert with parietal regions (Nyberg et al. 2003; Olesen et al. 2004). It is important to note that in contrast to studies focusing on learning, our study investigated training-induced changes in AV integration, not time-dependent neural modulations caused by learning per se. It is nevertheless possible that the observed inferior frontal as well as intraparietal activation increases might reflect consolidation and storage processes with regard to recently acquired crossmodal associations (see D'Esposito 2007 for a review).

Alternatively, it could be argued that increased inferior frontal and intraparietal activations reflect increased attention to artificial AV pairings after crossmodal association training (e.g., Corbetta and Shulman 2002; Duncan and Owen 2000).

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

Table 3
Summary of ROI-based analysis of the control contrasts

Cortical region	AV-ART-CON (POST) > AV-ART-INCON (PRE)	AV-ART-CON (POST) > AV-ART-CON (PRE)
Left IFC	$t(17) = -1.83$; $P = 0.085$	$t(17) = -1.53$; $P = 0.145$
Right pMTG	$t(17) = 1.333$; $P = 0.2$	$t(17) = 1.378$; $P = 0.186$
Right FG	$t(17) = -0.124$; $P = 0.903$	$t(17) = 0.257$; $P = 0.8$
Right STG	$t(17) = -0.293$; $P = 0.773$	$t(17) = -0.982$; $P = 0.340$

Note: Statistics were based on 2-tailed paired *t*-tests comparing the most repeated experimental condition (AV-ART-CON [POST]) to the 2 artificial AV conditions in the first experimental session. Repetition-related effects are most likely revealed by these contrasts. pMTG: posterior middle temporal gyrus.

There is solid evidence that novelty is one of the main triggers for an increase of attention-related activation (e.g., Herrmann and Knight 2001). Based on that, one would predict stronger fronto-parietal activity in the PRE session, where the artificial material was completely novel and unfamiliar. In contrast to this prediction, our results showed an extension of activation in the POST-training session, rendering the assumption unlikely that our findings merely reflect attention-related effects.

Another possible interpretation of our fronto-parietal findings could be that the AV integration-related cortical plasticity found here and the learning effects reported for similar regions in classic learning studies might reflect 2 complementary sides of the same coin. As predicted in our second hypothesis, the AV integration region in left aIFC was sensitive to (acquired) incongruency of artificial AV stimuli in the POST fMRI session, resembling the results for semantically incongruent natural object features in the PRE fMRI sessions. Taken together, these 2 findings indicate that AV integration in aIFC could serve the establishment and revision of AV associations, i.e., a function, which might also be part of the time-dependent activation changes in IFC found in the learning literature. In contrast, AV integration for artificial images and sounds in superior temporal cortex (STG/STS) was only found in the POST fMRI session, that is, when the formerly unfamiliar AV material has already been trained. However, these effects were not accompanied by AV-ART-POST > PRE effects. Moreover, STG/STS showed no sensitivity to the degree of semantic AV congruency, neither in the whole-brain congruency analysis (Fig. 4), nor when mapped based on the unimodal auditory (A-NAT vs. A-ART) contrast (Fig. 5B). This insensitivity to AV congruency suggests that STG/STS is rather involved in the integration of sounds and images, which are already familiar to a certain extent, than the establishment of novel AV associations. In this sense, IFC and STG/STS might have complementary roles in AV integration. Although STG/STS is integrating AV pairings that already reached a certain degree of familiarity, for example through training as in the present study, the IFC is either revising already known (but apparently incongruent) AV associations or establishing new ones. The explicit testing of this assumption would require information about the temporal flow of information. Although the sluggishness of the BOLD fMRI signal certainly forms a general limitation in this regard, fMRI-based measures of effective connectivity such as Granger causality mapping might be fruitfully applied in future studies.

In our current study, we investigated training-induced plasticity of AV integration using a passive paradigm which has been inspired by previous fMRI studies on AV integration (Calvert et al. 2000; Belardinelli et al. 2004; van Atteveldt et al. 2004, 2007). However, the potential impact of task demands on training-induced AV plasticity—especially in the aforementioned fronto-parietal regions—definitely forms another important issue to be investigated in future studies.

Training-Induced Congruency Effects

Besides incongruency effects in left anterior IFC, we found training-induced congruency effects for the POST fMRI session in the posterior aspect of right MTG and in right FG.

Our finding of an AV congruency effect in right pMTG is in line with the general involvement of posterior temporal regions in semantic processing. Significant activations of similar regions have been repeatedly shown particularly in the context of action- and tool-related processing (see e.g., Beauchamp et al. 2002; Johnson-Frey et al. 2005), not only with visual, but also with auditory presentations of the respective stimuli (see e.g., Lewis et al. 2005; Doehrmann et al. 2008). Interestingly, Murray et al. (2004) reported effects for repeatedly presented visual objects which had been paired previously with semantically congruent sounds. Source modeling suggested right-hemispheric effects in the proximity to the cluster reported here and close to lateral-occipital regions of higher-level visual cortex. Therefore, a critical role of pMTG in the context of conceptual processing has been suggested previously (see e.g., Martin 2007; Doehrmann and Naumer forthcoming). Our findings show that pMTG is also involved in crossmodal processing of objects which are more artificial than the ones investigated in previous studies (Beauchamp et al. 2002; Johnson-Frey et al. 2005) and should inspire further explorations of this region in future studies.

Ventral regions such as the FG have been repeatedly associated not only with perceptual processing of meaningful object stimuli, but also with the activation of conceptual representations in the course of thought and action (Martin and Chao 2001; Martin 2007; Doehrmann and Naumer forthcoming). Additionally, ventral temporal activation has also been found for AV stimulation with natural stimuli (Beauchamp et al. 2004; Amedi et al. 2005), and some degree of crossmodal plasticity in these regions has recently been demonstrated for higher-level stimuli such as faces and voices (von Kriegstein and Giraud 2006). Although our finding of stronger responses to congruent, that is, learnt, AV combinations is interesting from this perspective, we refrain from stronger interpretations of this finding in FG because, in contrast to our pMTG cluster, it was not accompanied by corresponding congruency effects for highly familiar natural stimuli. Certainly, further studies are needed to elucidate the functional role of these ventral temporal regions both for object-related AV processing in general and with respect to crossmodal plasticity in this domain.

Unlike in visual cortex, a unimodally defined auditory region in STG did not show any effect of semantic congruency for artificial AV stimuli. Interestingly, this region exhibited stronger responses to bimodal artificial conditions (compared with the respective unimodal conditions) only in the POST, but not in the PRE (training) session, but without showing any significant POST > PRE effect when directly comparing the respective artificial AV conditions. These findings are inconsistent to some extent with the results from our PRE fMRI session (Hein et al. 2007), which revealed stronger activation for semantically congruent than incongruent natural (i.e., animal) stimuli in

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

right STG. One possibility is that this inconsistency of results is driven by general differences between natural and artificial stimuli. Belardinelli et al. (2004) also reported a congruency effect in higher-level visual but not auditory regions, using some animal stimuli, but also images and sounds of tools. The latter might be more similar to our trained artificial AV objects, which might explain why a training-induced AV congruency effect was found in a similar region as reported by Belardinelli et al. (2004), instead of STG. Another possible explanation is that our training session was too short to induce a sufficient degree of familiarity. From the perspective of the results reported in this and our previous study (Hein et al. 2007) it is likely that prolonged and intensified training might not only enhance responses to bimodal compared with unimodal stimulation, but could also lead to the development of more pronounced congruency effects also in superior temporal regions. The extent of additional training with novel AV stimuli needed for the emergence of congruency effects similar to those for natural objects (Hein et al. 2007) or linguistic stimuli (van Atteveldt et al. 2004) remains to be determined.

In conclusion, our study demonstrates that even short-term crossmodal training of novel AV associations results in integration-related cortical plasticity and training-induced congruency effects for artificial AV stimuli in cortical regions especially of the frontal and (to a lesser degree) the temporal lobes, adding novel aspects to the understanding of object-related AV integration in the human brain.

Funding

German Ministry of Education and Research (BMBF); Frankfurt Medical School (Intramural Young Investigator Program) to M.J.N.; the Max Planck Society to L.M.; the Deutsche Forschungsgemeinschaft (HE 4566/1-1) to G.H.; and Society in Science/The Branco Weiss Foundation to G.H. Funding to pay the Open Access publication charges for this article was provided by the Intramural Young Investigator Program of Frankfurt Medical School.

Notes

We are grateful to Petra Janson for help with preparing the visual stimuli and Tim Wallenhorst for data acquisition support. *Conflict of Interest*: None declared.

Address correspondence to Marcus J. Naumer, PhD, Institut für Medizinische Psychologie, Goethe-Universität, Heinrich-Hoffmann-Strasse 10, D-60528 Frankfurt am Main, Germany. Email: M.J.Naumer@med.uni-frankfurt.de.

References

Altmann CF, Doehrmann O, Kaiser J. 2007. Selectivity for animal vocalizations in the human auditory cortex. Cereb Cortex. 17:2601–2608.

Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A. 2007. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. Nat Neurosci. 10:687.

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.

Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 41:809–823.

Beauchamp MS, Lee KE, Haxby JV, Martin A. 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron. 34:149–159.

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. Cognit Processing. 5:167.

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, Campbell R, Brammer MJ. 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the 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.

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

D'Esposito M. 2007. From cognitive to neural models of working memory. Philos Trans R Soc Lond B Biol Sci. 362:761–772.

Doehrmann O, Naumer MJ. Forthcoming. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res.

Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF. 2008. Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia. 46:2776–2786.

Duncan J, Owen AM. 2000. Common regions of 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.

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.

Hein G, Doehrmann O, Muller 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.

Herrmann CS, Knight RT. 2001. Mechanisms of human attention: event-related potentials and oscillations. Neurosci Biobehav Rev. 25:465–476.

James TW, Gauthier I. 2003. Auditory and action semantic features activate sensory-specific perceptual brain regions. Curr Biol. 13:1792–1796.

Johnson-Frey SH, Newman-Norlund R, Grafton ST. 2005. A distributed left hemisphere network active during planning of everyday tool use skills. Cereb Cortex. 15:681–695.

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.

Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25–45.

Martin A, Chao LL. 2001. Semantic memory and the brain: structure and processes. Curr Opin Neurobiol. 11:194–201.

Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A. 2004. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage. 21:125–135.

Nyberg L, Sandblom J, Jones S, Neely AS, Petersson KM, Ingvar M, Backman L. 2003. Neural correlates of training-related memory improvement in adulthood and aging. Proc Natl Acad Sci USA. 100:13728–13733.

Olesen PJ, Westerberg H, Klingberg T. 2004. Increased prefrontal and parietal activity after training of working memory. Nat Neurosci. 7:75–79.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020

Schwartz S, Maquet P, Frith C. 2002. Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. Proc Natl Acad Sci USA. 99:17137–17142.

Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.

Tanabe HC, Honda M, Sadato N. 2005. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J Neurosci. 25:6409–6418.

Taylor KI, Moss HE, Stamatakis EA, Tyler LK. 2006. Binding crossmodal object features in perirhinal cortex. Proc Natl Acad Sci USA. 103:8239–8244.

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 N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. Neuron. 43:271–282.

von Kriegstein K, Giraud A-L. 2006. Implicit multisensory associations influence voice recognition. PLoS Biol. 4:e326.

Weisberg J, van Turennout M, Martin A. 2007. A neural system for learning about object function. Cereb Cortex. 17:513–521.

Downloaded from https://academic.oup.com/cercor/article-abstract/19/7/1641/318107 by University of Glasgow user on 27 April 2020