◆ Human Brain Mapping 37:4333–4348 (2016) ◆

The Modulatory Effect of Semantic Familiarity on the Audiovisual Integration of Face-Name Pairs

Yuanqing Li,[1,2]* Fangyi Wang,[1,2] Biao Huang,[3] Wanqun Yang,[3] Tianyou Yu,[1,2] and Durk Talsma[4]

[1]*Center for Brain Computer Interfaces and Brain Information Processing, South China University of Technology, Guangzhou, China*
[2]*Guangzhou Key Laboratory of Brain Computer Interaction and Applications, Guangzhou, China*
[3]*Department of Radiology, Guangdong General Hospital, Guangzhou, China*
[4]*Department of Experimental Psychology, Ghent University, Ghent, Belgium*

Abstract: To recognize individuals, the brain often integrates audiovisual information from familiar or unfamiliar faces, voices, and auditory names. To date, the effects of the semantic familiarity of stimuli on audiovisual integration remain unknown. In this functional magnetic resonance imaging (fMRI) study, we used familiar/unfamiliar facial images, auditory names, and audiovisual face-name pairs as stimuli to determine the influence of semantic familiarity on audiovisual integration. First, we performed a general linear model analysis using fMRI data and found that audiovisual integration occurred for familiar congruent and unfamiliar face-name pairs but not for familiar incongruent pairs. Second, we decoded the familiarity categories of the stimuli (familiar vs. unfamiliar) from the fMRI data and calculated the reproducibility indices of the brain patterns that corresponded to familiar and unfamiliar stimuli. The decoding accuracy rate was significantly higher for familiar congruent versus unfamiliar face-name pairs (83.2%) than for familiar versus unfamiliar faces (63.9%) and for familiar versus unfamiliar names (60.4%). This increase in decoding accuracy was not observed for familiar incongruent versus unfamiliar pairs. Furthermore, compared with the brain patterns associated with facial images or auditory names, the reproducibility index was significantly improved for the brain patterns of familiar congruent face-name pairs but not those of familiar incongruent or unfamiliar pairs. Our results indicate the modulatory effect that semantic familiarity has on audiovisual integration. Specifically, neural representations were enhanced for familiar congruent face-name pairs compared with visual-only faces and auditory-only names, whereas this enhancement effect was not observed for familiar incongruent or unfamiliar pairs. *Hum Brain Mapp 37:4333–4348, 2016.* © 2016 Wiley Periodicals, Inc.

Key words: audiovisual face-name pairs; semantic familiarity; audiovisual semantic integration; decoding; reproducibility

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

Contract grant sponsor: National Key Basic Research Program of China (973 Program); Contract grant number: 2015CB351703; Contract grant sponsor: National High-tech R&D Program of China (863 Program); Contract grant number: 2012AA011601; Contract grant sponsor: National Natural Science Foundation of China; Contract grant numbers: 91420302, 81471654, and 61573150; Contract grant sponsor: Guangdong Natural Science Foundation; Contract grant number: 2014A030312005

Corrections added on 26 July 16, after online publication.
*Correspondence to: Yuanqing Li, Center for Brain Computer Interfaces and Brain Information Processing, South China University of Technology, Guangzhou 510640, China. E-mail: auyqli@scut.edu.cn
Received for publication 4 February 2016; Revised 25 June 2016; Accepted 1 July 2016.

DOI: 10.1002/hbm.23312
Published online 12 July 2016 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

The recognition of other individuals is a crucial component of social interaction. According to Bruce and Young [1986] information-processing model and its variants [Burton et al., 1990; Ellis et al., 1997; Stevenage et al., 2012; Valentine et al., 1991], this recognition might involve the processing of an individual's face, name and voice through separate yet parallel and interacting pathways [Belin et al., 2004; Blank et al., 2015; Campanella and Belin, 2007]. Face-voice integration arises from the interaction between the face and voice pathways, which have been the focus of many studies [Fairhall and Macaluso, 2009; Focker et al., 2011; Gonzalez-Castillo and Talavage, 2011; Joassin et al., 2004; Kamachi et al., 2003]. Face-voice integration might be partially segregated according to the type of information being integrated, e.g., speech information, affective information or identity information [Belin et al., 2004; Campanella and Belin, 2007]. Several studies have also addressed the neural mechanisms of visual face-name pair identification that arises from the interaction between the face and name pathways. For instance, both faces and names activate several brain areas, including the middle frontal lobe, middle parietal cortex (precuneus), and posterior cingulate cortex [Gorno-Tempini et al., 1998]. Furthermore, precuneus and middle frontal lobe activation is more extensive for familiar faces and names. The interaction between the face and name pathways might also involve visual faces and their corresponding spoken names. Audiovisual face-name pairs can appear in movies, television news, or social communication, and a name can be spoken by the individual with the corresponding face or by other individuals. To the best of our knowledge, however, no studies have investigated the audiovisual integration of face-name pairs, although it often occurs during person perception and social communication.

An audiovisual face-name pair can be familiar or unfamiliar. The mechanism through which semantic familiarity modulates the audiovisual integration of face-name pairs remains unclear. Several factors such as spatiotemporal contiguity [Pourtois and de Gelder, 2002; Stein and Meredith, 1993], crossmodal attention [Donohue et al., 2011; Koelewijn et al., 2010; Talsma et al., 2010], and the level of noise in audiovisual stimuli [Holmes, 2007] can influence audiovisual integration at the sensory or semantic levels. For instance, single-cell recordings [Meredith and Stein, 1983; Meredith and Stein, 1996] and functional magnetic resonance imaging (fMRI) experiments [Amedi et al., 2005; Calvert, 2001] have demonstrated that neural responses in the posterior superior temporal sulcus/middle temporal gyrus (pSTS/MTG) to audiovisual stimulation are the most pronounced for stimuli that coincide in space and time. More recent studies have indicated that semantic factors such as semantic congruence and stimulus familiarity can influence audiovisual integration [Doehrmann and Naumer, 2008; Yuval-Greenberg and Deouell, 2007]. For example, the semantic congruence of audiovisual stimuli increases activity in the lateral ventral and medial temporal cortex as well as the bilateral lingual gyrus, whereas semantic incongruence increases activity in regions of the left inferior frontal cortex [Belardinelli et al., 2004; Noppeney et al., 2008]. Hein et al. observed integration effects in the pSTS and superior temporal gyrus (STG) for highly familiar and semantically congruent audiovisual pairings (e.g., familiar animal sounds and images) and in the IFC for unfamiliar object images and sounds [Hein et al., 2007]. However, the effects of semantic factors on the neural correlates of audiovisual integration remain largely unknown. For audiovisual face-name pairs, we expect that semantic factors such as semantic familiarity will influence audiovisual integration; however, corresponding results have not been reported. In particular, research is needed regarding the effects of these semantic factors on the neural representations of audiovisual face-name pairs.

This study investigated the modulatory effects of semantic familiarity on the audiovisual integration of face-name pairs. According to the model of face and voice processing, integration involves both direct crosstalk between unimodal face or voice processing modules and interactions between unimodal face and voice regions and higher order, supramodal integration regions [Belin et al., 2004; Campanella and Belin, 2007]. This model suggests that information is transferred between these regions [Blank et al., 2011, 2015; Ethofer et al., 2012]. This neural mechanism might enhance the neural representations of the attended facial and vocal features (e.g., gender- or emotion-related features) of given audiovisual stimuli compared with V-fami-unfami and A-fami-unfami conditions, as we demonstrated previously [Li et al., 2015]. We expect that a similar neural mechanism underlies the audiovisual integration of face-name pairs and speculate that the differential integration of semantically familiar and unfamiliar face-name pairs occurs at the semantic level. Specifically, audiovisual semantic integration might enhance and enrich neural representations for semantically familiar congruent face-name pairs but not semantically familiar incongruent or unfamiliar pairs compared with visually presented faces or spoken names alone. To test this hypothesis, we conducted an fMRI experiment in which subjects were presented with visual-only facial images, auditory-only names or audiovisual face-name pairs that were semantically familiar and congruent, familiar and incongruent or unfamiliar and instructed to judge the familiarity of the stimuli (familiar vs. unfamiliar). We applied a multivariate pattern analysis (MVPA) to the fMRI data to directly assess the encoded semantic information related to the familiar feature, thereby observing the neural modulatory effects of the semantic familiarity of the stimuli upon the audiovisual integration of face-name pairs.

EXPERIMENTAL PROCEDURES AND METHODS

Subjects

Twelve healthy Chinese people (12 males; mean age ± SD, 29 ± 2 years old with normal or corrected-to-normal

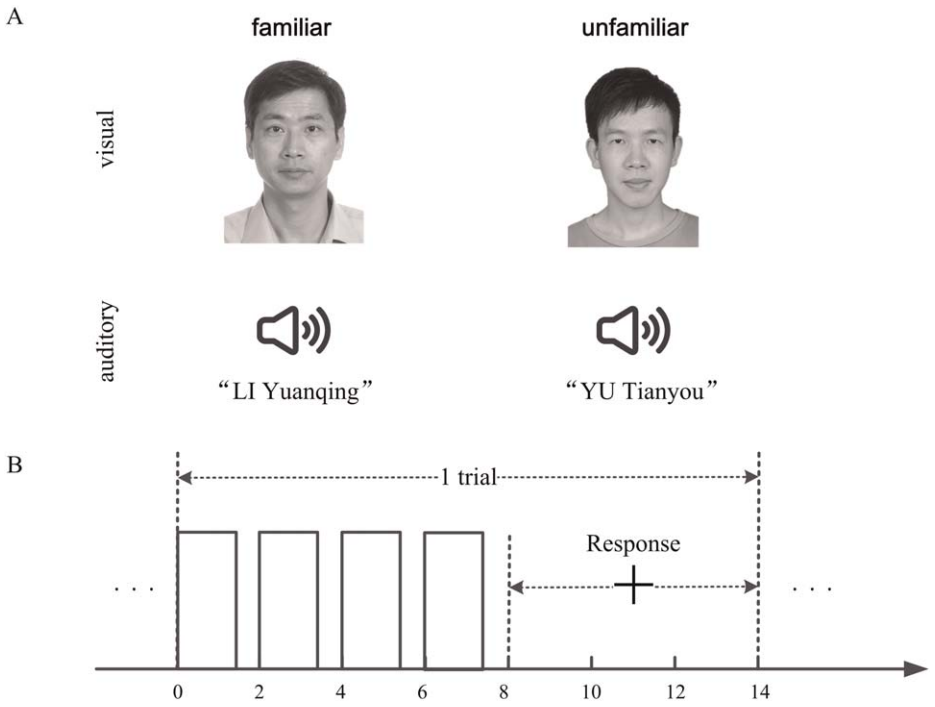

Figure 1.

(**A**) Two examples of the audiovisual face-name pairs. For a semantically familiar congruent/incongruent face-name pair, the face image and spoken name were matched/unmatched, whereas a semantically unfamiliar face-name pair was constructed using an unfamiliar face image and an unfamiliar spoken name from the Internet. (**B**) Trial time course. A stimulus was presented for 1.5 s and repeated four times during the initial 8 s of a trial. A fixation cross ("+") appeared on the 8th second, persisted for 6 s (response period), and changed colour on the 12th second. The subject made a familiarity judgement during the response period.

vision and normal hearing) participated in this study. All subjects provided written informed consent before the experiment. The Ethics Committee of Guangdong General Hospital, China approved the experimental protocol.

Experimental Stimuli and Design

For each subject, the visual stimuli included 45 images of familiar faces (friends, classmates, and relatives) and 45 images of unfamiliar faces (taken from the Internet). A woman spoke the auditory stimuli that included 45 semantically familiar Chinese names, each of which corresponded to a familiar face, and 45 semantically unfamiliar Chinese names derived from the Internet. Here, we prepared the auditory stimuli simulating the case: During face-name learning, the name corresponding to a face was often spoken by someone other than that individual. The subject provided the familiar facial images and names. The experimenter prepared the unfamiliar faces and names for each subject and tested the subject with these stimuli to ensure semantic unfamiliarity. We also processed these stimuli using Photoshop in the following way: Each image was converted to gray scale and adjusted to subtend 10.7° × 8.7° of the visual angle, and the luminance levels of the images were matched by adjusting the power value of each image (i.e., the sum of the squares of the pixel gray values; see examples in Fig. 1A). Each spoken name lasted approximately 1.5 s. The audio-power levels of these spoken names were matched by adjusting the total power value of each audio clip (Fig. 1A). Familiar audiovisual congruent face-name pairs were obtained by pairing the familiar facial images with their corresponding names, whereas the familiar audiovisual incongruent face-name pairs were generated by pairing familiar facial images with different names from other familiar individuals. To construct the unfamiliar audiovisual face-name pairs, we randomly paired each unfamiliar face with an unfamiliar name. In the following sections, the familiarity of the facial images, spoken names and face-name pairs as well as the audiovisual congruence of the familiar face-name pairs were considered. During the experiment, the visual stimuli were projected onto a screen using an LCD projector (SA-9900 fMRI Stimulation System, Shenzhen Sinorad Medical Electronics, Inc.), and the subjects viewed the visual

stimuli through a mirror mounted on a head coil. The auditory stimuli were delivered through a pneumatic headset (SA-9900 fMRI Stimulation System, Shenzhen Sinorad Medical Electronics, Inc.).

Each subject participated in four runs on a day, which were presented in a random order. Each run was composed of 70 trials (7 blocks of 10 trials), which corresponded to 70 visual-only, auditory-only, or audiovisual stimuli. Two runs were performed under visual-only and auditory-only conditions in which the visual-only stimuli included 35 familiar and 35 unfamiliar facial images (V-fami-unfami run), whereas the auditory-only stimuli included 35 familiar and 35 unfamiliar spoken names (A-fami-unfami run). The other two runs were conducted under audiovisual conditions. Specifically, the stimuli in an audiovisual run included 35 familiar congruent face-name pairs and 35 unfamiliar face-name pairs (AV-fami-cong-unfami run), whereas the stimuli in another audiovisual run contained 35 familiar incongruent face-name pairs and 35 unfamiliar face-name pairs (AV-fami-incong-unfami run), none of which overlapped with the face-name pairs presented in the AV-fami-cong-unfami run. Blank periods (gray screen with no auditory stimulation) of 20 s were placed between subsequent blocks. At the beginning of each run, five volumes were taken over a 10 s interval with no stimulation. The 70 stimuli were randomly assigned to the 70 trials, and the familiarity categories (familiar vs. unfamiliar) were balanced within each block. The subjects were asked to pay attention to the familiarity of the stimuli presented in each run (see Fig. 1A) and identify each stimulus as familiar or unfamiliar. Specifically, at the beginning of each block, a short instruction (e.g., "familiar 1 and unfamiliar 2" or "familiar 2 and unfamiliar 1") was displayed on the screen in Chinese for 4 s. For example, the instruction "familiar 1 and unfamiliar 2" instructed the subject to press key 1 for a familiar stimulus and key 2 for an unfamiliar stimulus. The two keys were pseudo-randomly assigned to the familiar and unfamiliar stimuli in each block. At the beginning of each trial, a visual-only, auditory-only, or audiovisual stimulus was presented for 1.5 s and then followed by a 0.5 s blank period. This 2 s cycle repeated four times with the same stimulus to enhance the signal-to-noise ratio of the fMRI responses and was followed by a 6 s blank period except that a fixation cross appeared on the screen to cue the subjects to press the keys according to the instruction for that block. For the AV-fami-incong-unfami run with familiar incongruent and unfamiliar face-name pairs, the subjects pressed the "familiar" key in response to familiar incongruent face-name pairs. The fixation cross changed colour on the 12th second to indicate that the next trial would begin shortly (2 s later, see Fig. 1B). Each run lasted 1,148 s. Note that we did not present familiar congruent, familiar incongruent or unfamiliar face-name pairs in the same audiovisual run; Otherwise, there would be 105 trials in the audiovisual condition, and the imbalance of the number of trials in the audiovisual condition, visual-only condition (70 trials), and auditory-only condition (70 trials) would not be useful for the comparison between these conditions.

fMRI Data Collection

The fMRI data were collected using a GE Signal Excite HD 3 T MR scanner at Guangdong General Hospital, China. A three-dimesnsional (3D) anatomical T1-weighted scan (FOV, 280 mm; matrix, 256 × 256; 128 slices; and slice thickness: 1.8 mm) was acquired before the functional scan for each subject each day. During the experiment, gradient-echo echo-planar (EPI) T2*-weighted images (25 slices acquired in an ascending noninterleaved order; TR = 2,000 ms, TE = 35 ms, flip angle = 70°; FOV: 280 mm, matrix: 64 × 64, slice thickness: 5.5 mm, no gap) were acquired, covering the entire brain.

Data Processing

Preprocessing

The fMRI data were preprocessed using SPM8 [Friston et al., 1994] and custom functions in MATLAB 7.4 (MathWorks, Natick, MA). Specifically, in each run, the first five volumes collected before magnetisation equilibrium was reached were discarded before the analysis. The following preprocessing steps were then performed on the fMRI data collected in each run: head-motion correction, slice-timing correction, co-registration between the functional and structural scans, normalization to the MNI standard brain, data masking to exclude nonbrain voxels, time-series detrending, and normalization of the time series in each block to a zero mean and one-unit variance.

General linear model (GLM) analysis

The fMRI experiment included four runs corresponding to the V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions, respectively. We performed a voxel-wise group analysis of the fMRI data based on a mixed-effect two-level GLM in SPM8 to determine whether audiovisual integration occurred for the familiar and unfamiliar stimuli in the AV-fami-cong-unfami and AV-fami-incong-unfami conditions. Specifically, the fMRI data from each subject were input into a first-level GLM, and the estimated beta coefficients across all subjects were then combined and analysed with a second-level GLM. We first compared the AV-fami-cong-unfami condition to the V-fami-unfami and A-fami-unfami conditions using the statistical criteria outlined below. Regarding the familiar congruent/unfamiliar stimuli, the following statistical criteria were used to identify the brain areas that exhibited audiovisual integration: [AV > max (A,V) ($P < 0.05$, FWE-corrected)] ∩ [$V > 0$ and $A > 0$ ($P < 0.05$, uncorrected)] [Beauchamp, 2005; Calvert and Thesen, 2004; Frassinetti et al.,

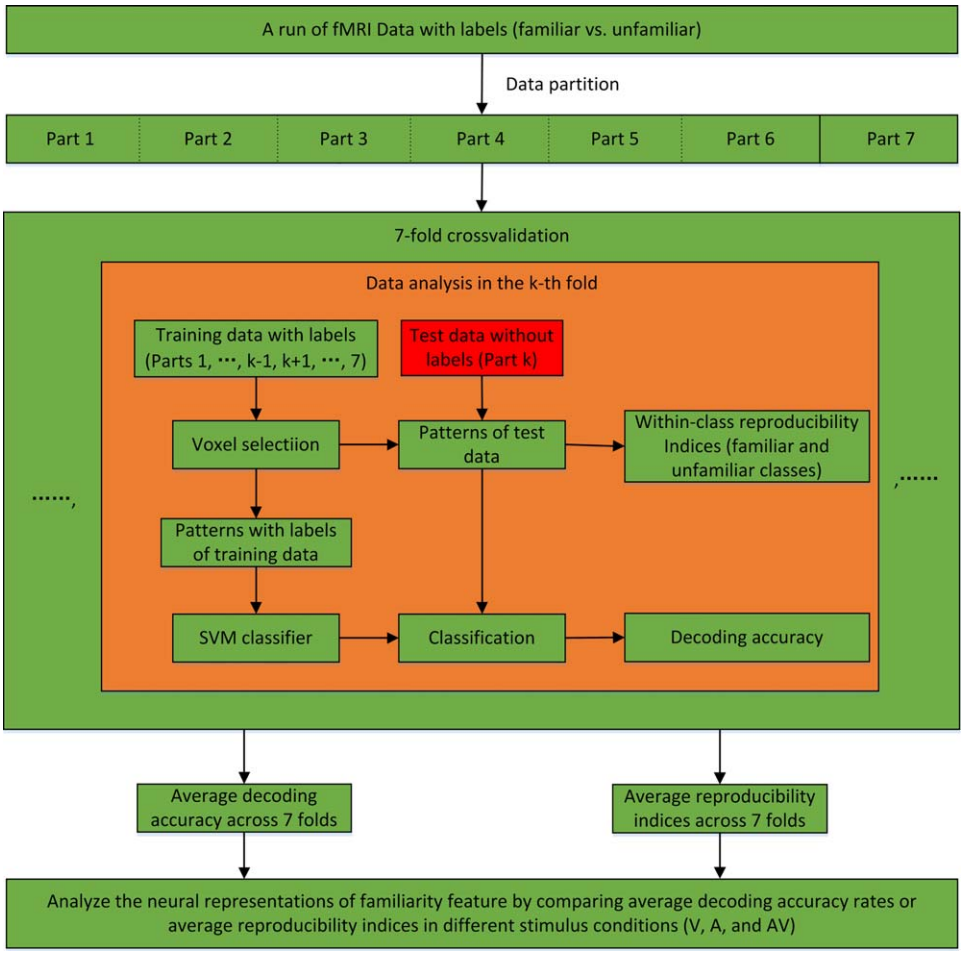

Figure 2.

MVPA procedure for the calculation of the reproducibility index and decoding accuracy in an experimental run. [Color figure can be viewed at wileyonlinelibrary.com]

2002; Macaluso and Driver, 2005], in which ∩ denotes the intersection of the two sets. For each subject, we also computed the percent signal changes in the identified pSTS/MTG clusters with a region-of-interest (ROI)-based analysis (in MATLAB toolbox MarsBaR-0.43 [Brett et al., 2002]). We identified the clusters that consisted of significantly activated voxels in the bilateral pSTS/MTG using the group GLM analysis described above. We also estimated the GLM model based on the mean BOLD signals of the clusters, and computed the percent signal change in each cluster as the ratio of the maximum of the estimated event response to the baseline. Next, we performed a similar analysis to compare the AV-fami-incong-unfami condition with the V-fami-unfami and A-fami-unfami conditions.

MVPA procedure

We performed an MVPA analysis of the fMRI data. The MVPA procedure was similar to that described in our previous study [Li et al., 2015]. Each subject performed four experimental runs: V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami. For each run, we first calculated two reproducibility indices of brain patterns corresponding to the familiar and unfamiliar categories by applying the MVPA method to the fMRI data. Higher reproducibility indicates stronger similarities within each class of brain patterns associated with the familiar or unfamiliar category. Using the fMRI data, we also decoded the familiarity categories (familiar vs. unfamiliar) of the stimuli for the subject. Below, we explain the MVPA procedure for each run.

The calculation of the reproducibility indices and the decoding for each experimental run/condition were performed through the sevenfold cross-validation illustrated in Figure 2. Specifically, the data from the 70 trials were evenly partitioned into seven nonoverlapping datasets. For the kth fold of the cross-validation ($k = 1, \ldots, 7$), the kth dataset (10 trials) was used for the test, and the remaining

six datasets (60 trials) were used for voxel selection and classifier training. Following the sevenfold cross-validation, the average reproducibility indices and the decoding accuracy rates were calculated across all folds. The data processing for the *k*th fold included the following steps. (1) Voxel selection. A spherical searchlight algorithm was applied to the training dataset for voxel selection [Kriegeskorte et al., 2006]. Specifically, this algorithm was sequentially centred at each voxel with a 3-mm radius searchlight that highlighted 19 voxels. Within each searchlight corresponding to a voxel, we computed a Fisher ratio through a Fisher linear discriminant analysis to indicate the level of discrimination between the two categories (familiar vs. unfamiliar) in the local neighbourhood of that voxel. In this way, a Fisher ratio map was obtained for the entire brain. We selected the K informative voxels with the highest Fisher ratios (e.g., $K = 1{,}600$ in this study). (2) fMRI activity pattern estimation. Using the selected voxels, a K-dimensional pattern vector was constructed for each trial in the training and test datasets. Specifically, because of the delayed hemodynamic response, we calculated each element of the pattern vector as the mean BOLD response of a selected voxel over 6 to 14 s of the trial (the last four volumes of each trial). (3) Reproducibility index calculation. We used $\cos\theta$ as a reproducibility index to assess the similarities in the fMRI activity patterns elicited by the stimuli, where θ is the angle between two pattern vectors, and larger $\cos\theta$ values indicate higher similarities. Specifically, we extracted 10 pattern vectors corresponding to the 10 trials of the test dataset, with five vectors in each class (familiar or unfamiliar stimuli). For each pair of pattern vectors within the same class, we calculated a reproducibility index. The mean of the reproducibility indices from each class was defined as a reproducibility index for the *k*th fold. Thus, two reproducibility indices were obtained for familiar and unfamiliar stimuli. (4) Decoding/prediction. To predict familiarity categories of the stimuli for the *k*th fold, a linear support vector machine (SVM) classifier was trained based on the pattern vectors of the labelled training data (+1 and −1 for the familiar and unfamiliar stimuli, respectively). The familiarity category of each trial of the test data was then predicted by applying the SVM to the corresponding pattern vector. After the seven-fold cross-validation, we obtained the decoding accuracy of each trial.

Localization of informative voxels

Using the data from the AV-fami-cong-unfami run, we obtained a voxel set, denoted V, which was informative for the two familiarity categories (familiar congruent vs. unfamiliar) as below. For each subject, we performed a sevenfold cross-validation for familiarity category decoding, as described previously. Based on the SVM training in each fold, we obtained an SVM weight map for the entire brain (the unselected voxels were assigned a weight of zero). The SVM weights reflected the importance of the voxels for decoding. By averaging the weight maps across all folds and all subjects, an actual group weight map was obtained for the differentiation of the familiarity categories. We then performed 1,000 permutations to obtain 1,000 group weight maps for the familiarity categories. Each group weight map was constructed as described above with the exception that, for each subject, the labels of all trials in the AV-fami-cong-unfami run were randomly assigned. To control the family-wise error (FWE) rate, a null distribution was constructed using the 1,000 maximum voxel weights, each of which was derived from a group weight map constructed in a permutation [Nichols and Hayasaka, 2003]. By thresholding the actual group weight map using the 95th percentile of the null distribution, we obtained the informative voxel set V.

RESULTS

Behavioral Results

All subjects successfully performed the discrimination task (familiar vs. unfamiliar) under the V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions. The average accuracy rates of all subjects for all conditions was >96%. The response accuracy did not significantly differ amongst the stimulus conditions ($P > 0.05$, one-way repeated-measures ANOVA with condition as the four-level factor [V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami]).

To examine the behavioral benefits of audiovisual integration, we conducted another behavioral experiment involving these 12 subjects using a procedure similar to that used in the fMRI experiment, except that the stimulus was presented only once for each trial, and the stimuli were presented with added Gaussian noise (signal noise rates: −6 dB for visual stimuli and −12 dB for auditory stimuli). The average accuracy rates with standard deviations were 0.849 ± 0.0579, 0.831 ± 0.053, 0.976 ± 0.0234, and 0.898 ± 0.081 for the V-fami-unfami condition, A-fami-unfami condition, AV-fami-cong-unfami condition, and AV-fami-incong-unfami condition, respectively. A one-way repeated-measures ANOVA with condition as the four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami) with regard to response accuracy indicated a significant main effect of stimulus condition ($P < 0.001$, $F_{(3,33)} = 8.387$). Furthermore, post hoc Bonferroni-corrected paired t-tests indicated that the accuracy rate was significantly higher for the AV-fami-cong-unfami condition than for the V-fami-unfami condition, A-fami-unfami condition, and AV-fami-incong-unfami conditions (all $P < 0.05$). No significant differences were found amongst the V-fami-unfami condition, A-fami-unfami condition, and AV-fami-incong-unfami conditions (all $P > 0.05$). The average response times with standard deviations were 1.691 ± 0.272, 2.188 ± 0.182, 1.255 ± 0.210, and 2.596 ± 0.166 s for the V-fami-unfami condition, A-fami-unfami condition, AV-fami-cong-unfami condition, and AV-fami-incong-unfami condition,

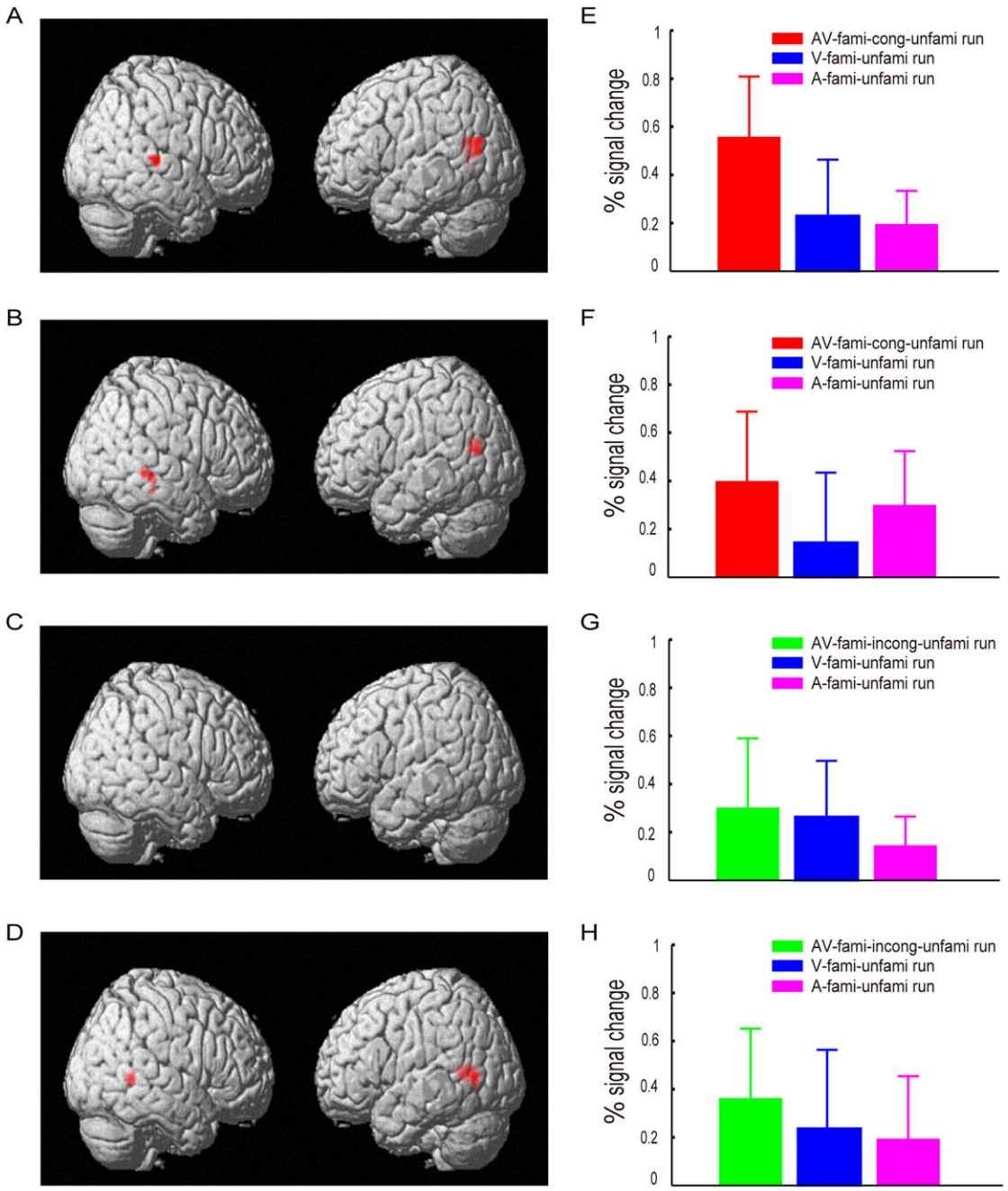

Figure 3.

Audiovisual integration in the brain areas that met the following criteria: [$AV > \max(A,V)$ ($P < 0.05$, FWE-corrected)] ∩ [$V > 0$ and $A > 0$ ($P < 0.05$, uncorrected)]. (**A**) Brain areas that exhibited the audiovisual integration of familiar congruent face-name pairs in the AV-fami-cong-unfami run. (**B**) Brain areas that exhibited the audiovisual integration of unfamiliar face-name pairs in the AV-fami-cong-unfami run. (**C**) No brain areas were identified to exhibit the audiovisual integration of familiar incongruent face-name pairs in the in the AV-fami-incong-unfami run. (**D**) Brain areas that exhibited audiovisual integration of unfamiliar face-name pairs in the AV-fami-incong-unfami run. These brain areas include left and right pSTS/MTG, with their coordinates shown in Table I. (**E**–**H**) Percent signal changes evoked by the audiovisual, visual-only and auditory-only stimuli in the bilateral pSTS/MTG activation clusters shown in A–D, respectively, where the percent signal changes in G were calculated using the activation clusters shown in A. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. The MNI coordinates of the clusters shown in Figure 3

Condition	Familiarity	Brain areas	MNI coordinates			mm^3
			X	*Y*	*Z*	
AV-fami-cong-unfami run	Familiar	Left pSTS/MTG	−48	−63	−3	3,240
		Right pSTS/MTG	57	−27	3	1,161
	Unfamiliar	Left pSTS/MTG	−54	−60	15	1,215
		Right pSTS/MTG	69	−33	−15	594
AV-fami-incong-unfami run	Familiar					
	Unfamiliar	Left pSTS/MTG	−48	−57	3	2,592
		Right pSTS/MTG	54	−48	3	1,377

respectively. The response time for each trial began at stimulus onset. A one-way repeated-measures ANOVA with condition as the four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami) with regard to response time indicated a significant main effect of stimulus condition ($P < 0.001$, $F_{(3,33)} = 74$). Furthermore, post hoc Bonferroni-corrected paired *t*-tests indicated that the response time was significantly lower for the AV-fami-cong-unfami condition compared with the V-fami-unfami condition, A-fami-unfami condition, and AV-fami-incong-unfami conditions (all $P < 0.05$). No significant differences were found amongst the V-fami-unfami, A-fami-unfami, and AV-fami-incong-unfami conditions (all $P > 0.05$).

Brain Areas Associated With Audiovisual Integration at the Sensory Level

Four experimental runs corresponded to the V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions (see Experimental procedures and methods). To determine whether audiovisual integration occurred, we performed a GLM analysis of the fMRI data at the group level and identified the heteromodal areas that exhibited enhanced neural responses in the audiovisual conditions (see Experimental procedures and methods). Specifically, we compared the AV-fami-cong-unfami condition with both the V-fami-unfami and A-fami-unfami conditions separately regarding the familiar and unfamiliar stimuli using the following criteria: [AV > max (A,V) ($P < 0.05$, FWE-corrected)] ∩ [$V > 0$ and $A > 0$ ($P < 0.05$, uncorrected)]. Similar comparisons were also performed for the AV-fami-incong-unfami condition and the V-fami-unfami and A-fami-unfami conditions. Figure 3 and Table I show audiovisual integration of the familiar audiovisual congruent and unfamiliar audiovisual face-name pairs but not for the familiar incongruent face-name pairs.

Decoding Results

For each experimental run of each subject, we separately decoded the familiarity categories (i.e., familiar and unfamiliar) of the stimuli from the collected fMRI data using the MVPA method (see Experimental procedures and methods). We systematically varied the number of selected voxels from 25 to 3,000 to decode the familiarity categories, and the average decoding accuracy rates of all subjects are shown in Figure 4A. With different numbers of selected voxels for decoding, we were able to obtain different accuracy rates. Generally, we needed to determine an optimal/suboptimal number of voxels based on the training data to obtain a satisfactory decoding result. As Figure 4A shows, through the consideration of varying numbers of voxels, we easily observed that the decoding accuracies were higher for the AV-fami-cong-unfami condition than for the V-fami-unfami, A-fami-unfami, or AV-fami-incong-unfami conditions. Thus, we did not need to determine an optimal/suboptimal number of selected voxels for decoding, which simplified our data analysis. As an example, we used 1,600 selected voxels to present the decoding and statistical results. The decoding results obtained from these 1,600 selected voxels are shown in Figure 4B. Furthermore, a one-way repeated-measures ANOVA with condition as a four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions) indicated a significant main effect of stimulus condition ($P < 10^{-6}$, $F_{(3,33)} = 18.67$). Post hoc Bonferroni-corrected paired *t*-tests regarding stimulus condition indicated that the decoding accuracy was significantly higher for the AV-fami-cong-unfami condition ($83.2 \pm 2.3\%$) than for the V-fami-unfami ($63.9 \pm 3.5\%$), A-fami-unfami ($60.4 \pm 2.3\%$), or AV-fami-incong-unfami conditions ($67.9 \pm 1.2\%$; all $P < 0.005$, corrected). No significant differences were observed amongst the V-fami-unfami, A-fami-unfami, and AV-fami-incong-unfami conditions (all $P > 0.05$). According to Figure 4A, the decoding accuracy rates for each condition did not significantly vary for numbers of selected voxels larger than 1,600. Thus, we were able to use any number between 1,600 and 3,000 of selected voxels to obtain similar comparison results.

Although the voxels selected based on the training data were informative, a small number of voxels (e.g., 50 voxels) were not sufficient for effective decoding (see Fig. 4). In this case, the decoding accuracy significantly improved

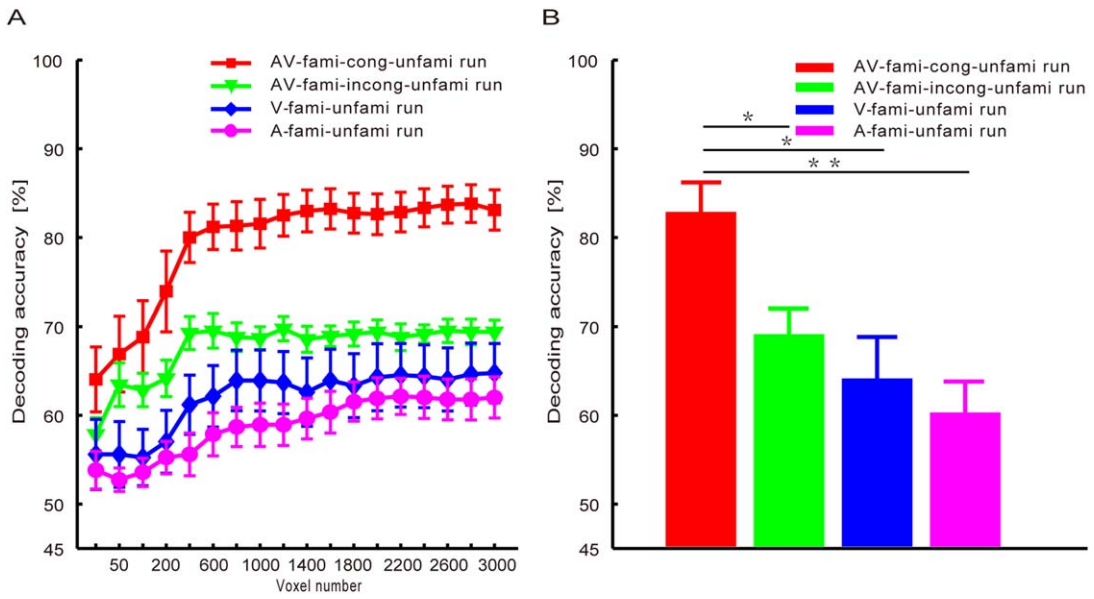

Figure 4.

The average decoding accuracies of the familiar vs. unfamiliar stimuli across all subjects for the V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions. (**A**) Decoding accuracy curves with respect to the numbers of selected voxels. (**B**) Decoding accuracy rates obtained with 1,600 selected voxels. *$P < 0.005$, corrected; **$P < 0.001$, corrected. [Color figure can be viewed at wileyonlinelibrary.com]

after adding informative voxels. For instance, a significant increase in decoding accuracy occurred after 50 voxels was augmented to 600 voxels (Fig. 4A). However, after the number of voxels increased to 1,600 (Fig. 4A), the decoding accuracy did not significantly vary with respect to the number of voxels. In this case, more voxels might represent redundant information.

Reproducibility

Using the MVPA method, we calculated two reproducibility curves with respect to the number of selected voxels (from 25 to 3,000), corresponding to the familiar and unfamiliar categories within each experimental run (see Experimental procedures and methods). These curves are shown in Figure 5A,C. As an example, Figure 5B,D show the reproducibility indices obtained using 1,600 selected voxels for familiar and unfamiliar stimuli. As Figure 5A shows, we easily observed that the reproducibility indices of the familiar stimuli were significantly higher for the audiovisual congruent face-name pairs than for the audiovisual incongruent face-name pairs, facial images, or auditory names after considering varying numbers of voxels, whereas the reproducibility indices of the familiar stimuli did not significantly differ amongst the AV-fami-incong-unfami, V-fami-unfami, and A-fami-unfami conditions. It follows from Figure 5C that the reproducibility indices of the unfamiliar stimuli did not significantly differ amongst the AV-fami-cong-unfami, AV-fami-incong-unfami, V-fami-unfami, and A-fami-unfami conditions.

In the following analysis, we used 1,600 selected voxels as an example to present the statistical results. We performed a two-way repeated-measures ANOVA with condition as a four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions) and stimulus familiarity as a two-level factor (familiar and unfamiliar) for the reproducibility indices. Significant main effects of stimulus condition ($P < 10^{-4}$, $F_{(3,33)} = 9.68$) and stimulus familiarity ($P < 0.001$, $F_{(1,11)} = 20.66$) were observed (Fig. 5B,D). In addition, a significant interaction between stimulus condition and familiarity was found ($P < 10^{-6}$, $F_{(3,33)} = 18.33$). A one-way repeated measures ANOVA with condition as a four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions) indicated a significant main effect of stimulus condition ($P < 10^{-6}$, $F_{(3,33)} = 18.1$; Fig. 5B) with regard to the reproducibility indices for familiar stimuli. Furthermore, post hoc Bonferroni-corrected paired *t*-tests indicated that the reproducibility index was significantly higher for familiar audiovisual congruent stimuli than for familiar visual-only, auditory-only, or audiovisual incongruent stimuli (all $P < 0.005$, corrected). No significant differences were observed amongst familiar visual-only, auditory-only, and audiovisual incongruent stimuli (all $P > 0.05$). In addition, a one-way repeated-measures ANOVA with condition as a four-level factor (V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami

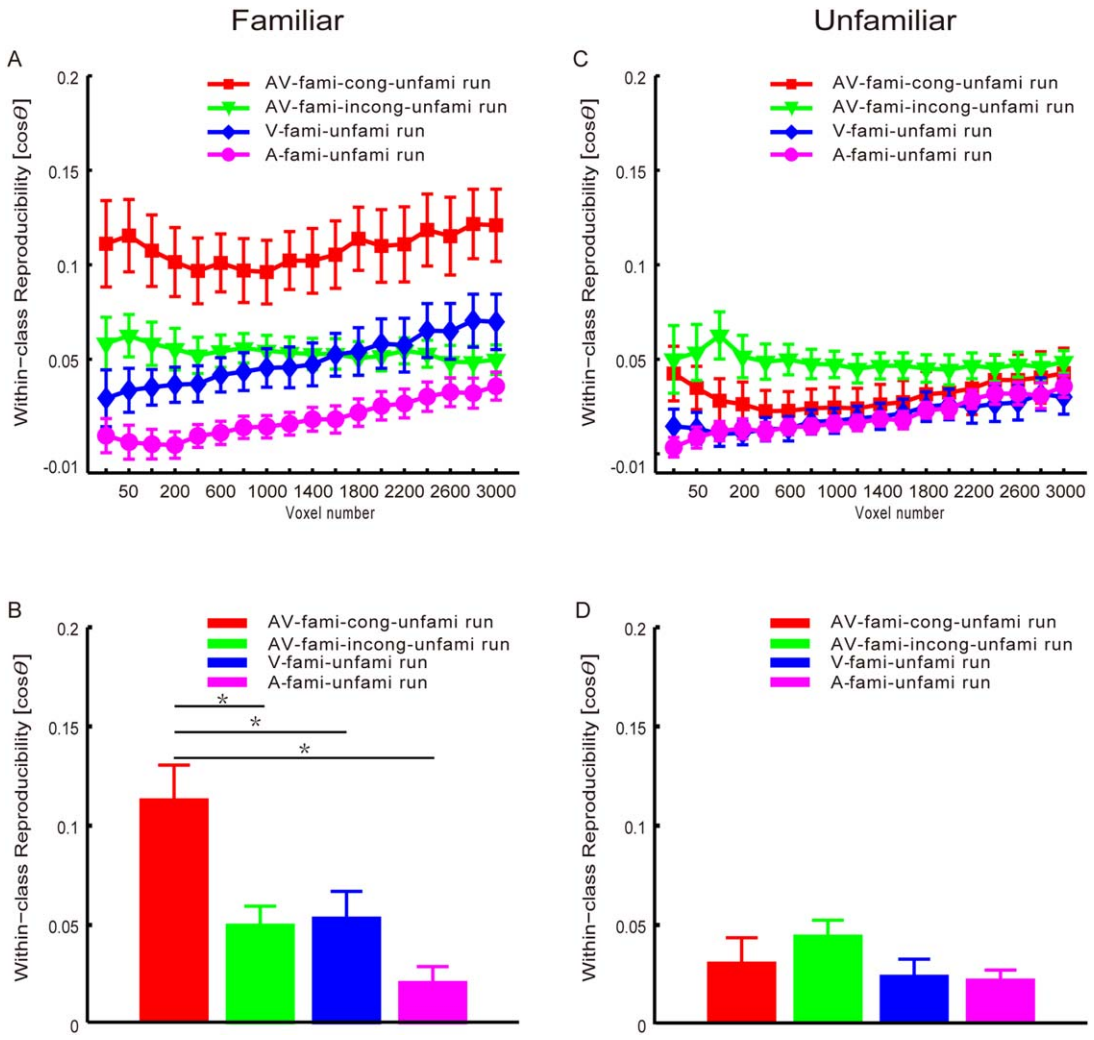

Figure 5.

Reproducibility indices (means and standard errors across all subjects). First row: the reproducibility curves with respect to the numbers of selected voxels for the AV-fami-cong-unfami, AV-fami-incong-unfami, V-fami-unfami, and A-fami-unfami runs. Second row: the reproducibility results obtained with 1,600 selected voxels. The asterisks indicate significant differences ($P < 0.005$, corrected). Left: familiar stimuli (facial images in the V-fami-unfami, names in the A-fami-unfami, congruent face-name pairs in the AV-fami-cong-unfami, and incongruent face-name pairs in the AV-fami-incong-unfami runs). Right: unfamiliar stimuli (facial images in the V-fami-unfami, names in the A-fami-unfami, face-name pairs in the AV-fami-cong-unfami, and face-name pairs in the AV-fami-incong-unfami runs). We used as a reproducibility index to assess the similarity of two fMRI activity patterns elicited by the stimuli, where θ is the angle between the two pattern vectors. [Color figure can be viewed at wileyonlinelibrary.com]

conditions) did not indicate a significant main effect of stimulus condition ($P = 0.12$, $F_{(3,33)} = 2.13$; Fig. 5D) regarding the reproducibility indices of the unfamiliar stimuli.

Informative Voxels

Using the data collected in the AV-fami-cong-unfami condition, we obtained voxels that were informative for discriminating familiarity (see Experimental procedures). The distribution of these informative voxels is shown in Table II.

DISCUSSION

This study investigated the neural modulatory effects of semantic familiarity on the audiovisual integration of face-name pairs. In the fMRI experiment, semantic familiar/unfamiliar stimuli were presented under visual-only (facial images), auditory-only (spoken names), and audiovisual conditions (familiar congruent, familiar incongruent or unfamiliar face-name pairs), and the subjects reported the stimuli as either familiar or unfamiliar. To assess the

TABLE II. Distribution of voxels that were informative for discrimination (familiar vs. unfamiliar). These voxels were obtained using data from the AV-fami-cong-unfami run ($P < 0.05$, corrected)

		MNI coordinates				
Brain region	Side	*X*	*Y*	*Z*	Weight	mm^3
Cuneus	L	−6	−63	27	0.071	594
	R	6	−87	15	0.070	432
Supplementary motor area	L	−3	0	57	0.071	432
	R	15	9	60	0.056	594
Precentral gyrus	R	42	−9	60	0.078	432
Superior frontal gyrus	L	−15	60	18	0.072	756
	R	15	27	60	0.047	972
Middle frontal gyrus	R	33	12	60	0.062	405
Inferior frontal gyrus, triangular part	R	45	30	12	0.105	459
Inferior frontal gyrus, orbital part	R	54	42	−6	0.083	540
Fusiform gyrus	L	−27	−33	−15	0.060	459
Angular gyrus	L	−45	−54	30	0.052	432
	R	33	−51	39	0.077	594
Superior frontal gyrus, medial	L	−9	54	30	0.091	1539
	R	6	60	6	0.081	621
Superior temporal gyrus	L	−42	−18	−3	0.081	1404
	R	60	−12	6	0.086	2538
Middle temporal gyrus	L	−54	−6	−21	0.075	1377
	R	51	−33	0	0.090	432
Lingual gyrus	L	−9	−57	3	0.117	837
	R	6	−69	0	0.096	1107
Superior frontal gyrus, medial orbital	L	0	45	−12	0.086	513
	R	9	69	−15	0.088	405
Insula	R	42	−6	6	0.056	486
Anterior cingulate gyrus	L	−6	54	0	0.061	459
Median cingulate gyrus	L	−6	−42	42	0.075	513
Posterior cingulate gyrus	L	−3	−39	24	0.091	432
Calcarine fissure and surrounding cortex	L	−6	−60	9	0.152	1458
	R	6	−57	12	0.137	945
Superior occipital gyrus	R	27	−81	39	0.105	891
Middle occipital gyrus	L	−36	−84	3	0.059	405
Precuneus	L	−6	−57	12	0.122	1890
	R	12	−51	42	0.098	1053
Inferior temporal gyrus	L	−54	−63	−6	0.051	432

semantic information related to familiarity encoded in the brain, we decoded the familiarity categories and calculated the reproducibility indices of brain patterns for semantically familiar and unfamiliar stimuli in each stimulus condition using an MVPA on the fMRI data.

Behavioral Results

The behavioral results in terms of response accuracy revealed no significant differences amongst the V-fami-unfami, A-fami-unfami, AV-fami-cong-unfami, and AV-fami-incong-unfami conditions, likely because each stimulus was presented four times, the subjects had sufficient time to make a judgement, and all of the tasks were easy. Although no behavioral benefits were observed under any audiovisual condition, the fMRI results indicated that audiovisual integration occurred for familiar congruent and unfamiliar face-name pairs (Fig. 3). Notably, although the task was relatively easy, the fMRI results discussed below showed that the neural representations of familiar face-name pairs were processed more efficiently than those of any other face-name pairs. In the behavioral experiment, we added noise to the visual and auditory stimuli and reduced the number of stimulus repetitions in each trial. Under those conditions, the behavioral benefits of audiovisual integration became apparent.

Sensory-Level Audiovisual Integration of the Familiar Congruent and Unfamiliar Face-Name Pairs

The neural mechanisms of audiovisual integration have previously been investigated using neuroimaging techniques, and several brain regions including the pSTS/MTG have been identified as heteromodal areas [Bushara et al., 2003; Calvert et al., 2000; Frassinetti et al., 2002]. Specifically, increased neural activity has been observed in the

pSTS/MTG when audiovisual congruent stimuli were compared with visual-only and auditory-only stimuli. By contrast, enhanced neural activity in the pSTS/MTG might indicate the occurrence of audiovisual integration [Bushara et al., 2003; Calvert et al., 2000; Calvert and Thesen, 2004; Frassinetti et al., 2002; Macaluso and Driver, 2005]. Furthermore, the pSTS/MTG generally acts as a presemantic, heteromodal region for processing crossmodal perceptual features [Taylor et al., 2006b); that is, the audiovisual integration in the pSTS/MTG generally occurs at the sensory level. In our experiment, we observed increased fMRI activity in the pSTS/MTG for both familiar congruent and unfamiliar face-name pairs (Fig. 3A,B,D–F,H), suggesting that audiovisual integration occurs at the sensory level. However, no increased fMRI activity in the pSTS/MTG was observed for familiar incongruent face-name pairs (Fig. 3C,G), suggesting that no audiovisual integration occurred.

Modulatory Effects of the Semantic Familiarity of Stimuli on the Audiovisual Integration of Face-Name Pairs

We showed that familiar congruent face-name pairs modulated neural activities. The results shown in Figure 4 indicate that the decoding accuracy rate calculated from the fMRI data was significantly higher in the AV-fami-cong-unfami condition than in the visual-fami-unfami and A-fami-unfami conditions. This increased decoding accuracy was not observed in the AV-fami-incong-unfami condition. Furthermore, our results showed a significantly improved reproducibility index for the brain patterns of familiar congruent face-name pairs compared with the brain patterns of facial images or auditory names alone. Reproducibility was not improved for the brain patterns of familiar incongruent or unfamiliar face-name pairs (Fig. 5). In the audiovisual conditions, audiovisual integration occurred for familiar congruent and unfamiliar face-name pairs but not familiar incongruent face-name pairs (Fig. 3). Together, the results indicate different effects of audiovisual integration for familiar congruent and unfamiliar face-name pairs. That is, the audiovisual integration of the familiar congruent face-name pairs but not the unfamiliar face-name pairs improved the reproducibility of the corresponding brain patterns and resulted in improved differentiation between the two classes of brain patterns corresponding to the familiar congruent and unfamiliar categories. This result might be because the audiovisual integration of familiar congruent face-name pairs occurred at both the sensory and semantic levels, whereas the audiovisual integration of unfamiliar face-name pairs occurred only at the sensory level. Overall, neural representations in the brain were improved only for the familiar audiovisual congruent stimuli compared with the visual-only and auditory-only stimuli.

In this study, we used both the decoding accuracy and the reproducibility index to assess the neural representations of stimuli. The decoding involved the discrimination between two classes of brain patterns that corresponded to familiar vs. unfamiliar stimuli. For a class of brain patterns, a larger average reproducibility index indicates the higher similarity of these brain patterns and implies more consistent fMRI data, which is useful for decoding. Thus, the decoding accuracy and the reproducibility index are associated with each other. However, an obvious difference exists between decoding accuracy and the reproducibility index. Specifically, the decoding accuracy cannot depict the characteristics of each class of brain patterns; however, the reproducibility indices may reflect different similarities for two classes of brain patterns. In this study, our results indicated that the decoding accuracy was higher for the AV-fami-cong-unfami condition than for the V-fami-unfami, A-fami-unfami, or AV-fami-incong-unfami condition. Furthermore, a comparison of the AV-fami-cong-unfami condition with the V-fami-unfami, A-fami-unfami and AV-fami-incong-unfami conditions revealed that the reproducibility of brain patterns was enhanced for familiar congruent face-name pairs. The reproducibility indices of the brain patterns associated with the unfamiliar stimuli did not significant differ amongst these experimental conditions. Therefore, the high decoding accuracy in the AV-fami-cong-unfami condition primarily originated from the enhanced reproducibility of the brain patterns associated with the familiar congruent face-name pairs and not the unfamiliar pairs.

Brain Areas Informative for Discriminating Familiarity in the Audiovisual Condition With Familiar Congruent and Unfamiliar Face-Name Pairs

Using the data collected during the AV-fami-cong-unfami run, we localised the voxels that were informative for decoding familiarity, which were distributed across different brain areas (Table II). Several brain regions including the left fusiform gyrus (FG), bilateral STG, bilateral MTG, left posterior cingulate gyrus, bilateral precuneus, right medial frontal gyrus, and right inferior frontal gyrus were involved in decoding the familiarity category.

Our results are partially consistent with previous evidence related to facial processing, as described below. First, functional neuroimaging studies have identified several brain regions in the occipital and temporal areas, including the FG, inferior occipital gyrus, and STS, which are active during the perception of familiar and unfamiliar faces [Halgren et al., 1999; Hoffman and Haxby, 2000]. Furthermore, both faces and names activate several brain areas including the middle frontal lobe and precuneus spreading to the posterior cingulate cortex [Gorno-Tempini et al., 1998]. Second, in addition to common face-selective regions, familiar faces can activate the amygdala,

hypothalamus, posterior cingulate cortex, medial frontal cortex, and right hippocampus [Leveroni et al., 2000; Pierce et al., 2004; Sugiura et al., 2001], which are associated with representations of semantic information about an individual (e.g., name, occupation, interests, and place of origin; [Gobbini and Haxby, 2007]. In addition, familiarity, regardless of modality (i.e., face or voice), activates areas in the paracingulate gyrus, which is suggested to be a multimodal "familiarity-checking" processor [Shah et al., 2001]. The increased activity in the bilateral posterior cingulate gyrus has been attributed to increases in familiarity with faces [Kosaka et al., 2003]. Third, in each trial of the AV-fami-cong-unfami run, the repeated presentation of unfamiliar stimuli might have been involved in the formation of new face-name associations. The learning of new visual face-name associations is supported by a distributed network of brain regions that includes the orbital frontal gyrus and medial frontal gyrus [Sperling et al., 2001]. These brain areas were also activated in our experiment (Table II), although we used audiovisual face-name pairs. The integration of unfamiliar sounds and images and familiar incongruent materials involves the inferior frontal regions, whereas the integration of familiar sounds and images is correlated with pSTS activation [Hein et al., 2007]. Furthermore, audiovisual integration in the inferior frontal regions might subserve the learning of associations between AV object features that, once learned, are integrated in the pSTS [Hein et al., 2007]. In our experiment, the pSTS was also involved in audiovisual integration at the sensory level regarding the unfamiliar face-name pairs, possibly because the degree of familiarity was much higher for unfamiliar face-name pairs than for the unfamiliar artificial stimuli used by Hein et al. [2007].

Our results could be incorporated into the Interactive Activation and Competition (IAC) model of personal recognition [Burton et al., 1990]. Bruce and Young developed an information-processing model [Bruce and Young, 1986] in which the recognition of a familiar face is achieved through the initial structural processing of the face and subsequent stages of identity-information and name retrieval. Burton et al. extended this model and proposed the IAC model of personal recognition, which includes face-recognition units (FRUs), name input units, personal identity nodes (PINs), and semantic information units. An individual's face, voice, or written or spoken name might activate PINs. The IAC model was extended, and several variants were proposed by incorporating name [Valentine et al., 1991] and voice recognition [Ellis et al., 1997; Stevenage et al., 2012]. It is common in these models that face, name and voice processing are attributed to separate yet parallel pathways with interactions and information transfers between these pathways [Belin et al., 2004; Blank et al., 2015; Campanella and Belin, 2007]. The interactions between the face and voice pathways might be reflected during the audiovisual integration of faces and voices. Furthermore, two neural mechanisms underlie the audiovisual integration of faces and voices [Campanella and Belin, 2007]. One such mechanism is the recruitment of supramodal convergence regions, including the bilateral posterior STS regions that are most likely involved in general audiovisual integration, and additional regions such as the amygdaloid complex for affective information as well as the precuneus/retrosplenial cortex and anterior temporal lobe regions for identity information. Another mechanism is the multimodal influence on "unimodal" processing stages, which are implemented via direct anatomo-functional coupling between unimodal cortical processing modules and/or feedback projections from the heteromodal cortex. The current study considered audiovisual person recognition based on face-name pairs and explored the neural correlates of the corresponding audiovisual semantic integration. Our results might be explained within the person-recognition framework described above. First, we observed that neural representations of familiar congruent face-name pairs but not familiar incongruent and unfamiliar pairs were enhanced and enriched compared with those of visual-only faces or auditory-only names. This enhancement effect might be the result of interactions between the face and name pathways in the brain. We previously revealed similar findings for faces and voices [Li et al., 2015]; that is, the audiovisual integration of the face and voice enhanced the neural representations of features of an individual that were selectively attended to (such as gender- or emotion-related features).

Familiar faces are represented by rich visual semantic and emotional codes that support nearly effortless perception and recognition. Our superior ability to recognize familiar faces compared with unfamiliar faces most likely stems from differences in the quality and richness of our neural representations of these faces [Natu and O'Toole, 2011]. This phenomenon might also apply to familiar names. Furthermore, for a semantically familiar and congruent audiovisual face-name pair, the association has been established throughout an individual's learning history [Doehrmann and Naumer, 2008]. These rich neural representations and associations might underlie the effective audiovisual semantic integration of the familiar and congruent face-name pairs demonstrated in this study. When people are presented with unfamiliar audiovisual face-name pairs, an association between the visual and auditory stimuli is gradually established through learning. In this case, audiovisual interactions at the sensory level also occur (see Fig. 3); however, audiovisual integration is not effectively implemented at the semantic level because of the absence of semantic information or the association between visual and auditory stimuli.

Effective Decoding of Familiarity of Face-Name Pairs

Our results showed that the average decoding accuracy was $83.2 \pm 2.3\%$ in the AV-fami-cong-unfami condition (Fig. 4). Human faces share many common features. The

overlap of features makes it difficult to discriminate between the different semantic categories related to faces (e.g., old vs. young, male vs. female, and so on) using fMRI signals [Taylor et al., 2006a]. For instance, [Haxby et al., 2011] presented a high-dimensional model of the representational space in the human ventral temporal (VT) cortex, and the response-pattern vectors (measured using fMRI) from individuals' voxel spaces were mapped onto this common model space. The results showed that, based on fMRI signals, the classifier distinguished human faces from nonhuman animal faces as well as monkey faces from dog faces but not human female from human male faces. Axelrod and Yovel systematically explored the role of face-selective areas in the recognition of famous faces using an MVPA and showed that the fusiform face area (FFA) was the only region where face identities could be discriminated based on multivoxel patterns [Axelrod and Yovel, 2015]. An average decoding accuracy of approximately 60% was obtained when discriminating two famous faces. Using fMRI and an MVPA, Anzellotti et al. investigated where tolerance across viewpoints is achieved in the human ventral stream and showed that occipitotemporal cortex and anterior temporal lobe do not merely represent specific images of faces; but also the identity information with tolerance for image transformations [Anzellotti et al., 2013]. In their analyses, linear SVM classifiers were trained to discriminate two face identities, and the decoding accuracy rates significantly higher than chance level (50%) were achieved; however, they ranged from 55% to 57%. Our high decoding accuracy rate associated with the AV-fami-cong-unfami condition most likely did not originate from double dipping in the pattern analysis because of the following reasons. (i) Double dipping in the pattern analysis generally occurs when an overlap exists between the training and test data. In our MVPA, the training data, which were used for feature selection and classifier training, and the test data for each fold of cross-validation were nonoverlapping (see Fig. 2). This setting avoided double dipping. (ii) The decoding accuracy rates obtained using the same MVPA were not high for the V-fami-unfami and A-fami-unfami conditions ($63.9 \pm 3.5\%$ and $60.4 \pm 2.3\%$, respectively). In addition, we performed 20 permutation tests. In each permutation, we applied the same MVPA to the fMRI data from the AV-fami-cong-unfami condition with the randomly assigned labels. The obtained average decoding accuracy rate ($52.06 \pm 2.59\%$) was not significant higher than chance (i.e., 50%; $P > 0.05$). These results demonstrated that double dipping did not exist with regard to our MVPA. Otherwise, we would have obtained high decoding accuracy rates in the V-fami-unfami and A-fami-unfami conditions as well as in the 20 permutations using the same MVPA.

In each experimental run/condition of the current study, the subjects were instructed to judge the familiarity categories of the stimuli (familiar vs. unfamiliar) whilst the fMRI data were collected. Our MVPA of each experimental run decoded the familiarity categories of the stimuli from the collected fMRI data. Some association might exist between the difficulty of the experimental task and the decoding accuracy. For instance, if discriminating two categories is difficult behaviorally, then the corresponding decoding accuracy based on the fMRI data is generally low. According to our behavioral results, subjects found it significantly easier to distinguish familiar congruent face-name pairs from unfamiliar face-name pairs compared with all other conditions. Our MVPA also obtained the highest decoding accuracy for the AV-fami-cong-unfami run. Of course, it is not easy to establish an explicit relationship between the difficulty of an experimental task and its corresponding decoding accuracy. However, our MVPA did not decode the difficulty of the familiarisation task because we did not distinguish or compare two experimental conditions of different difficulties in the separate decoding performed for each experimental run/condition. Furthermore, we calculated the average percent signal changes at informative or randomly selected voxels and compared these average percent signal changes across different experimental conditions. The results showed that the improvement of the decoding accuracies in the AV-fami-cong-unfami condition, compared with the other conditions, was not the result of general differences of overall activity amongst these experimental conditions (Supporting Information).

CONCLUSIONS

This study explored the modulatory effect of the semantic familiarity of face-name pairs on audiovisual integration. Our GLM analysis indicated that audiovisual integration occurred for both familiar congruent and unfamiliar face-name pairs but not for familiar incongruent face-name pairs. Furthermore, an MVPA revealed that the neural representations of semantically familiar congruent face-name pairs were enhanced compared with visual-only facial images and auditory-only names. This enhancement effect was not observed for unfamiliar or familiar incongruent pairs. In the Supporting Information section, we show that this modulatory effect of semantic familiarity with regard to congruent stimuli in audiovisual integration might have arisen from an enhanced functional connectivity that influences information flow from the heteromodal bilateral perirhinal cortex to the brain areas that encode familiarity. In the future, we will consider other types of audiovisual stimuli and extensively explore semantic information exchange across visual and auditory modalities to test and expand our conclusions.

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.

Anzellotti S, Fairhall SL, Caramazza A (2013): Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988–1995.
Axelrod V, Yovel G (2015): Successful decoding of famous faces in the fusiform face area. PLoS One 10:e0117126.
Beauchamp MS (2005): Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3:93–113.
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.
Belin P, Fecteau S, Bedard C (2004): Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.
Blank H, Anwander A, von Kriegstein K (2011): Direct structural connections between voice- and face-recognition areas. J Neurosci 31:12906–12915.
Blank H, Kiebel SJ, von Kriegstein K (2015): How the human brain exchanges information across sensory modalities to recognize other people. Hum Brain Mapp 36:324–339.
Brett M, Anton JL, Valabregue R, Poline JB (2002): Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage 16:S497.
Bruce V, Young A (1986): Understanding face recognition. Br J Psychol 77:305–327.
Burton AM, Bruce V, Johnston RA (1990): Understanding face recognition with an interactive activation model. Br J Psychol 81:361–380.
Bushara KO, Hanakawa T, Immisch I, Toma K, Kansaku K, Hallett M (2003): Neural correlates of cross-modal binding. Nat Neurosci 6:190–195.
Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123.
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, Thesen T (2004): Multisensory integration: Methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191–205.
Campanella S, Belin P (2007): Integrating face and voice in person perception. Trends Cogn Sci 11:535–543.
Doehrmann O, Naumer MJ (2008): Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. Brain Res 1242:136–150.
Donohue SE, Roberts KC, Grent-'t-Jong T, Woldorff MG (2011): The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. J Neurosci 31:7982–7990.
Ellis HD, Jones DM, Mosdell N (1997): Intra- and inter-modal repetition priming of familiar faces and voices. Br J Psychol 88:143–156.
Ethofer T, Bretscher J, Gschwind M, Kreifelts B, Wildgruber D, Vuilleumier P (2012): Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb Cortex 22:191–200.
Fairhall SL, Macaluso E (2009): Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. Eur J Neurosci 29:1247–1257.
Focker J, Holig C, Best A, Roder B (2011): Crossmodal interaction of facial and vocal person identity information: An event-related potential study. Brain Res 1385:229–245.
Frassinetti F, Bolognini N, Ladavas E (2002): Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res 147:332–343.
Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1994): Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp 2:189–210.
Gobbini MI, Haxby JV (2007): Neural systems for recognition of familiar faces. Neuropsychologia 45:32–41.
Gonzalez-Castillo J, Talavage TM (2011): Reproducibility of fMRI activations associated with auditory sentence comprehension. Neuroimage 54:2138–2155.
Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, Frackowiak RS (1998): The neural systems sustaining face and proper-name processing. Brain 121: 2103–2118.
Halgren E, Dale AM, Sereno MI, Tootell RB, Marinkovic K, Rosen BR (1999): Location of human face-selective cortex with respect to retinotopic areas. Hum Brain Mapp 7:29–37.
Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ (2011): A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72:404–416.
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.
Hoffman EA, Haxby JV (2000): Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3:80–84.
Holmes NP (2007): The law of inverse effectiveness in neurons and behavior: multisensory integration versus normal variability. Neuropsychologia 45:3340–3345.
Joassin F, Maurage P, Bruyer R, Crommelinck M, Campanella S (2004): When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. Neurosci Lett 369:132–137.
Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003): 'Putting the face to the voice': Matching identity across modality. Curr Biol 13:1709–1714.
Koelewijn T, Bronkhorst A, Theeuwes J (2010): Attention and the multiple stages of multisensory integration: A review of audiovisual studies. Acta Psychol 134:372–384.
Kosaka H, Omori M, Iidaka T, Murata T, Shimoyama T, Okada T, Sadato N, Yonekura Y, Wada Y (2003): Neural substrates participating in acquisition of facial familiarity: an fMRI study. Neuroimage 20:1734–1742.
Kriegeskorte N, Goebel R, Bandettini P (2006): Information-based functional brain mapping. Proc Natl Acad Sci USA 103: 3863–3868.
Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000): Neural systems underlying the recognition of familiar and newly learned faces. J Neurosci 20:878–886.
Li Y, Long J, Huang B, Yu T, Wu W, Liu Y, Liang C, Sun P (2015): Crossmodal integration enhances neural representation of task-relevant features in audiovisual face perception. Cereb Cortex 25:384–395.
Macaluso E, Driver J (2005): Multisensory spatial interactions: A window onto functional integration in the human brain. Trends Neurosci 28:264–271.
Meredith MA, Stein BE (1983): Interactions among converging sensory inputs in the superior colliculus. Science 221: 389–391.
Meredith MA, Stein BE (1996): Spatial determinants of multisensory integration in cat superior colliculus neurons. J Neurophysiol 75:1843–1857.

Natu V, O'Toole AJ (2011): The neural processing of familiar and unfamiliar faces: A review and synopsis. Br J Psychol 102: 726–747.

Nichols T, Hayasaka S (2003): Controlling the familywise error rate in functional neuroimaging: A comparative review. Stat Methods Med Res 12:419–446.

Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008): The effect of prior visual information on recognition of speech and sounds. Cereb Cortex 18:598–609.

Pierce K, Haist F, Sedaghat F, Courchesne E (2004): The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. Brain 127:2703–2716.

Pourtois G, de Gelder B (2002): Semantic factors influence multisensory pairing: A transcranial magnetic stimulation study. Neuroreport 13:1567–1573.

Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001): The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. Brain 124:804–815.

Sperling RA, Bates JF, Cocchiarella AJ, Schacter DL, Rosen BR, Albert MS (2001): Encoding novel face-name associations: a functional MRI study. Hum Brain Mapp 14:129–139.

Stein BE, Meredith MA (1993): The Merging of the Senses. Cambridge, MA: The MIT Press.

Stevenage SV, Hugill AR, Lewis HG (2012): Integrating voice recognition into models of person perception. J Cogn Psychol 24: 409–419.

Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, Ito K, Fukuda H (2001): Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. Neuroimage 13:877–890.

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010): The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14:400–410.

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

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

Valentine T, Bredart S, Lawson R, Ward G (1991): What's in a name? Access to information from people's names. Eur J Cogn Psychol 3:147–176.

Yuval-Greenberg S, Deouell LY (2007): What you see is not (always) what you hear: Induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27:1090–1096.