

Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex

Citation for published version (APA):

van Atteveldt, N. M., Formisano, E., Goebel, R. W., & Blomert, L. P. M. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage*, 36(4), 1345-1360. <https://doi.org/10.1016/j.neuroimage.2007.03.065>

Document status and date:

Published: 01/01/2007

DOI:

[10.1016/j.neuroimage.2007.03.065](https://doi.org/10.1016/j.neuroimage.2007.03.065)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex

Nienke M. van Atteveldt,* Elia Formisano, Rainer Goebel, and Leo Blomert

University of Maastricht, Faculty of Psychology, Department of Cognitive Neuroscience, P.O. Box 616, 6200 MD Maastricht, The Netherlands
Maastricht Brain Imaging Center (M-BIC), Maastricht, The Netherlands

Received 19 February 2007; revised 20 March 2007; accepted 24 March 2007
Available online 18 April 2007

In alphabetic scripts, letters and speech sounds are the basic elements of correspondence between spoken and written language. In two previous fMRI studies, we showed that the response to speech sounds in the auditory association cortex was enhanced by congruent letters and suppressed by incongruent letters. Interestingly, temporal synchrony was critical for this congruency effect to occur. We interpreted these results as a neural correlate of letter-sound integration, driven by the learned congruency of letter-sound pairs. The present event-related fMRI study was designed to address two questions that could not directly be addressed in the previous studies, due to their passive nature and blocked design. Specifically: (1) to examine whether the enhancement/suppression of auditory cortex are truly multisensory integration effects or can be explained by different attention levels during congruent/incongruent blocks, and (2) to examine the effect of top-down task demands on the neural integration of letter-sound pairs. Firstly, we replicated the previous results with random stimulus presentation, which rules out an explanation of the congruency effect in auditory cortex solely in terms of attention. Secondly, we showed that the effects of congruency and temporal asynchrony in the auditory association cortex were absent during active matching. This indicates that multisensory responses in the auditory association cortex heavily depend on task demands. Without task instructions, the auditory cortex is modulated to favor the processing of congruent and synchronous information. This modulation is overruled during explicit matching when all audiovisual stimuli are equally relevant, independent of congruency and temporal relation.

© 2007 Elsevier Inc. All rights reserved.

Introduction

In speech-based alphabetic scripts, letters and speech sounds are the basic elements of correspondence between written and spoken language. Investigations of the mechanism by which letters and

speech sounds are associated in the brain are therefore important for a better understanding of the neural basis of literacy. In two previous fMRI studies using blocked stimulus presentation and passive perception (Van Atteveldt et al., 2004, 2007), we showed that the response to speech sounds in the auditory association cortex is enhanced by congruent letters and suppressed by incongruent letters (hereafter referred to as “congruency effect”). Interestingly, temporal synchrony was critical for this congruency effect to occur. Furthermore, we revealed an enhanced response to bimodally presented letter-sound pairs relative to their unisensory components in heteromodal regions in the superior temporal sulcus/gyrus (STS/STG). This response pattern suggested an integrative role for the STS/STG, which conformed to other reports providing evidence for its importance in multisensory integration of identity information (Calvert, 2001; Amedi et al., 2005; Beauchamp, 2005). Since the auditory association cortex did not respond to letters alone, we interpreted the observed congruency effect as a feedback effect from the STS/STG. This is in accordance with other studies advocating a feedback mechanism for audiovisual integration of speech information (Calvert et al., 2000; King and Calvert, 2001). There were no explicit task demands in these two previous studies, so these findings most likely reflect a default or automatic integration mechanism for the over-learned associations between letters and speech sounds.

Due to the passive nature of the previous studies and the blocked stimulus presentation, several questions could not directly be addressed. First of all, the congruency effect in the auditory cortex might be explained partly in terms of attention or arousal because of the blocked stimulus presentation (see e.g. Beauchamp, 2005). We interpreted the congruency effect as an effect of multisensory integration, possibly mediated through STS/STG. However, in a blocked design, it may be sufficient to integrate only the first cross-modal stimulus of a block because all subsequent stimuli have the same congruency. After the first stimulus, attention levels may be increased or decreased depending on the congruency. Although the specificity of the congruency effect makes it unlikely to be a general effect of attention, this explanation can only fully be excluded if the congruency effect is replicated with random stimulus presentation.

A second open question concerns the effect of different task instructions on the proposed default integration mechanism. Since

* Corresponding author. University of Maastricht, Faculty of Psychology, Department of Cognitive Neuroscience, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Fax: +31 43 3884125.

E-mail address: N.vanAtteveldt@psychology.unimaas.nl
(N.M. van Atteveldt).

Available online on ScienceDirect (www.sciencedirect.com).

there were no explicit task demands in the two previous studies, no top-down task-related influences were exerted over the integration process. Therefore, the incoming information was probably integrated based on stimulus-related properties only, i.e. content congruency and temporal coincidence. However, next to stimulus-related binding factors, top-down factors generated by task instructions also seem to have a clear effect on multisensory integration (Andersen et al., 2004). This is especially relevant in the context of letter–sound integration, since the use of the learned letter–sound associations is likely to be functional in different reading and spelling activities instead of being restricted to passive perception. A more specific question is whether task instruction affects the strong temporal constraints found for passive integration (Van Atteveldt et al., 2007). We predict that if the bimodal information is made functionally relevant independent of temporal coincidence, neural integration effects will be found over a broader temporal window.

In the present fMRI study, we aimed to address these two open issues by examining the effects of stimulus presentation mode (blocked vs. random) and task instruction (passive perception vs. active matching) on the neural mechanism for letter–speech sound integration. We focused on the congruency effect in the auditory association cortex, because it was demonstrated to be a robust finding and exhibited a strong effect of temporal asynchrony. Furthermore, as already discussed in Van Atteveldt et al. (2007), the congruency effect can be used as an indicator of multisensory integration, since a distinction between corresponding and non-corresponding letters and speech sounds cannot be established unless the unisensory inputs have been integrated successfully. However, it cannot be excluded that a congruency effect may reflect integration only indirectly, i.e. through feedback from another brain region (e.g. STS/STG). Therefore, regions exhibiting a congruency effect are not necessarily performing integrative operations themselves. Nevertheless, whether direct or indirect, an effect of congruency reliably indicates that integration has been performed. To examine the effect of task instruction, we employed a same/different matching task because it forces the integration process to be explicit while leaving stimulus presentation unchanged (congruent and incongruent letter–sound pairs). This task therefore enables a direct comparison of the congruency effect during passive and active integration of letters and speech sounds with different stimulus onset asynchronies (SOAs).

Both research aims require random stimulus presentation, and thereby the switch from a blocked to an event-related fMRI design, which has consequences that might confound the aimed results. In both designs, we presented the stimuli in silent intervals between subsequent volume scans (Jäncke et al., 2002). Since single stimuli were presented in the event-related design, the silent intervals were shorter than in the blocked design, which relatively increased the amount of scanner noise. Therefore, the hemodynamic response in auditory areas may be more saturated, and the range in which the fMRI signal can increase due to experimental auditory stimulation may be limited in the event-related design. A second, more general consequence of using event-related fMRI is that statistical power to detect experimental effects is reduced, since fMRI responses to single stimuli (compared to blocks of stimuli) are used to identify differences between conditions (Friston et al., 1998). Although it has reliably been shown that single events can elicit detectable signal changes (e.g. Buckner et al., 1996), it is not clear how the reduction in stimulation duration will affect the relative signal changes to single letters and speech sounds in an event-related fMRI design.

The present report consists of two different studies. In Study 1, each subject participated in three fMRI experiments: a passive blocked, a passive event-related, and an active event-related experiment in which the matching task was introduced. The intermediate (passive event-related) design enabled direct comparison of blocked and random stimulus presentation in a passive setting. This had two goals: (1) testing the hypothesis that the congruency effect found in the previous studies might be explained by attention or arousal due to the blocked presentation, and (2) controlling for more general effects of switching from blocked to event-related fMRI unrelated to task effects (higher scanner noise and shorter stimulus duration). After accounting for the general effects of using event-related fMRI in Study 1, we investigated the effect of task instruction on letter–speech sound integration at different SOAs in Study 2. We first selected suitable SOAs in a behavioral experiment. Then, subjects in Study 2 participated in two fMRI experiments: a passive blocked and an active event-related experiment in which the matching task was performed on letter–sound pairs presented at different SOAs.

Materials and methods

Stimuli

The same stimulus material was used in Studies 1 and 2. Stimuli were speech sounds corresponding to single letters and their visually presented counterparts (vowels a, e, i, y, o, u, consonants d, g, h, k, l, n, p, r, s, t, z). Letters and sounds were presented separately/unimodally or in bimodal matching (congruent, e.g. “a” – /a/) or non-matching (incongruent, e.g. “a” – /e/) combinations. Bimodal combinations were either both vowels or both consonants. Speech sounds were digitally recorded (sampling rate 44.1 kHz, 16-bit quantization) from a female native Dutch speaker and represented isolated speech sounds (phonemes) and not letter names. The selected speech sounds were recognized 100% correct in a pilot experiment ($n=10$). Recordings were band-pass filtered (180–10,000 Hz) and resampled at 22.05 kHz. Average duration of the speech sounds was 352 ms (± 5 ms), the average sound intensity level was approximately 70 dB SPL. White lower case letters (typeface “Arial”) were presented for 350 ms on a black background. During the fMRI experiments, the visual stimuli were projected (Sanyo PLC-XT16) onto a frosted screen positioned at the rear end of the MR scanner bore, and viewed by the participants through a mirror mounted onto the head coil. Auditory stimuli were presented with a MR-compatible Intercom Commander XG MRI Audio System from Resonance Technologies Inc. The sound from this system is transduced by a 2-way stereo headset that also serves as ear defender. The headset is air tube driven for the lower frequencies in combination with a non-magnetic piezo tweeter. Stimulus presentation was synchronized with the scanner pulses using the software package “Presentation” (<http://nbs.neuro-bs.com>). Participants’ responses were registered by a handheld fiber-optic response system (LUMItouch fMRI Optical Response keypad, Photon Control, Burnaby, Canada, www.photonixco.com).

Study 1

Participants

Twelve healthy native Dutch subjects (4 males, mean age 23, range 20–27) participated in the first fMRI experiment. All subjects were university students enrolled in an undergraduate study program and were selected based on a questionnaire including questions

Table 1
Schematic overview of the systematically manipulated factors in studies 1 and 2

Manipulation	Study			
	Study 1–exp1 Study 2–exp1	Study 1–exp2	Study 1–exp3	Study 2–exp2
fMRI design/ stimulus presentation	Blocked	Event-related/ random	Event-related/ random	Event-related/ random
Task	Passive perception	Passive perception	Active matching	Active matching
Letter-sound SOA (ms)	0 (synchronous)	0 (synchronous)	0 (synchronous)	0 or +/-300

concerning present or past reading or other language problems. All were right-handed, had normal or corrected-to-normal vision, and normal hearing capacity. Subjects gave informed written consent and were paid for their participation.

Experimental procedure

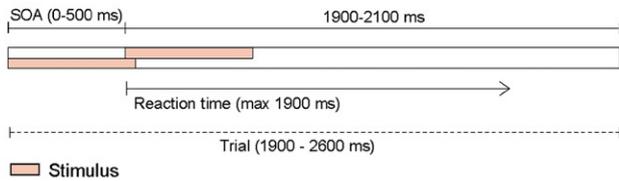
All subjects participated in 3 different fMRI experiments (see Table 1): a passive experiment in which the stimuli were presented

in blocks (passive blocked), a passive experiment in which the stimuli were presented randomly (passive event-related) and an active experiment in which the stimuli were presented randomly and a response was required to each stimulus (active event-related).

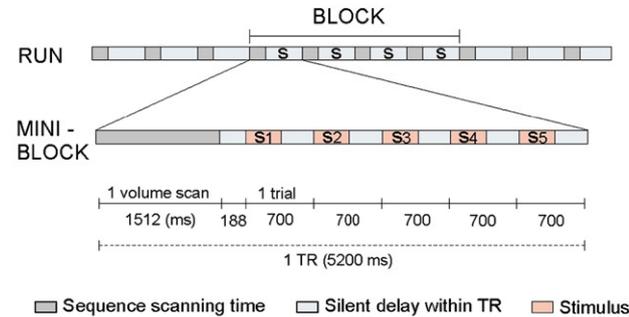
In the passive blocked runs (Fig. 1B) stimuli were presented in blocks of four different conditions: unimodal visual, unimodal auditory, bimodal congruent and bimodal incongruent. In the bimodal conditions, visual letters and speech sounds were presented simultaneously. Experimental blocks (20.8 s) were composed of 4 mini-blocks of 5.2 s (=1 sequence repeat time (TR), see Scanning procedure). In the first 1512 ms of each mini-block, one brain volume was acquired. No stimuli were presented in this period, only a fixation cross to keep the eyes of the subjects focused on the center of the screen. In the subsequent silent 3688 ms, 5 stimuli were presented. In each run, 2 blocks (of 20 stimuli in total) were presented of each condition. Fixation periods were presented in the beginning and end of each run (26 s), and between each experimental block (20.8 s). The order of the experimental blocks was pseudo-randomized within runs and counterbalanced across runs.

In the passive event-related runs (Fig. 1C, Study 1), experimental trials of 2 s (=1 TR) consisted of single stimuli of the same four conditions (unimodal visual, unimodal auditory, bimodal congruent

A behavioral experiment



B fMRI experiments - Block runs



C fMRI experiments - Event-related runs

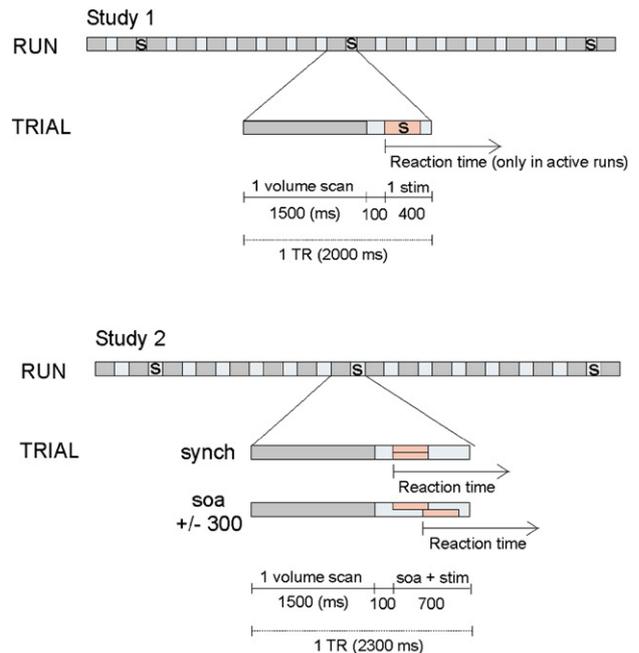


Fig. 1. Schematic description of the different experimental designs used in Studies 1 and 2. (A) Trials in the behavioral experiment. At the beginning of each trial, a visual and an auditory stimulus were presented at different SOAs (0–500 ms in both directions). Reaction times were measured from the onset of the second stimulus. (B) Blocked design used in fMRI experiments in Studies 1 and 2. Effective TR was 5200 ms, silent delay within each TR was 3688 ms. Experimental blocks were composed of 4 mini-blocks of 1 TR. Each mini-block started with the acquisition of one volume scan followed by 5 stimuli (S1–S5) presented in the silent delay period. (C) Event-related design used in fMRI experiments in Studies 1 (upper figure) and 2 (lower figure). Study 1: Effective TR was 2000 ms, silent delay within each TR was 500 ms. Experimental trials (every 7th, 8th or 9th TR) started with the acquisition of one volume scan followed by 1 stimulus (S) presented in the silent delay period. Reaction times (only in the active runs) were measured from the onset of the stimulus pair. Study 2: Effective TR was 2300 ms, silent delay within each TR was 800 ms. Experimental trials (every 6th, 7th or 8th TR) started with the acquisition of one volume scan followed by 1 stimulus (S) presented in the silent delay period. The stimuli were presented at different SOAs: –300 (visual first), 0 (synch=synchronous) or 300 (auditory first) ms. Reaction times were measured from the onset of the second stimulus. SOA=stimulus onset asynchrony, TR=sequence repeat time (sequence scanning time + silent delay), Stim=stimulus.

and bimodal incongruent). In the bimodal trials, letters and speech sounds were always presented simultaneously. In the first 1.5 s of each trial, one brain volume was acquired. In the subsequent silent 0.5 s, 1 stimulus was presented. The stimuli were presented randomly with a jittered intertrial interval (ITI) of 14, 16 or 18 s (corresponding to 7, 8, or 9 TRs). A fixation cross was presented on the screen between the stimuli. In each run, 10 trials of each condition were presented.

In the active event-related runs (Fig. 1C, Study 1), the same design was used as in the passive event-related runs, except that only bimodal stimuli were presented (congruent and incongruent). Again, letters and speech sounds were always presented simultaneously. In each run, 20 trials of each condition were presented. Subjects were instructed to respond as fast and accurately as possible to the question: “Are the letter and the sound the same?” after each stimulus pair. Reaction times were recorded from the onset of the stimulus pair. The correct button (middle or index finger of the right hand) for “yes” and “no” was counterbalanced over subjects.

Scanning procedure

In two of the subjects, imaging was performed on a 3-T whole-body system (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany) located at the F.C. Donders center in Nijmegen, The Netherlands. In ten of the subjects, imaging was performed on a 3-T head scanner (Magnetom Allegra, Siemens Medical Systems, Erlangen, Germany) located at the Maastricht Brain Imaging Centre (M-BIC) in Maastricht, The Netherlands. In each subject, 2–4 passive blocked runs of 70 volumes were acquired, followed by 2–4 passive event-related runs of 330 volumes, followed by 2 active event-related runs of 330 volumes. The passive runs were always acquired before the active runs to ensure that passive perception was uncontaminated with the explicit matching instruction. In four of the subjects, all runs were acquired in the same session. In eight of the subjects, the passive blocked runs were acquired in an earlier session.

A BOLD-sensitive EPI sequence was used for all functional scans (matrix 64×64 , 24 slices, slice thickness 4 or 4.5 mm, FoV 192 or 224 mm²; both depending on the subject's head size), resulting voxel size $3 \times 3 \times 4$ or $3.5 \times 3.5 \times 4.5$ mm³, TE/TR slice 30/63 ms (blocked runs) and 30/62.5 (event-related runs), FA 90°. As schematically shown in Figs. 1B–C, the acquisition of one whole-brain volume was clustered in the beginning of each TR. This provides a silent delay within each TR in which stimulus perception is uncontaminated by EPI noise, and is demonstrated to be highly efficient for studying the auditory cortex with fMRI (Jäncke et al., 2002; Van Atteveldt et al., 2004, 2007). Due to using a silent delay, the effective TR (sequence repeat time) was longer than the sequence scanning time. In the passive blocked runs, the effective TR was 5200 ms, sequence scanning time was 1512 ms, delay within a TR was 3688 ms. During experimental blocks, 5 stimuli were presented in this delay (see Fig. 1B). In the event-related runs, the effective TR was 2000 ms, sequence scanning time was 1500 ms, delay within a TR was 500 ms. During experimental trials, one stimulus was presented in this delay (see Fig. 1C, Study 1). A high-resolution structural scan (voxel size: $1 \times 1 \times 1$ mm³) was collected for each subject using a T1-weighted 3D MP-RAGE sequence (Magnetization-Prepared Rapid Acquisition Gradient Echo, TR = 2.3 s, TE = 3.93 ms, 192 sagittal slices).

Analysis of fMRI time-series

Functional and anatomical images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands;

Goebel et al., 2006). The following preprocessing steps were performed: slice scan time correction (using sinc interpolation), linear trend removal, temporal high pass filtering to remove low-frequency nonlinear drifts of 3 or less cycles per time-course, and 3D motion correction to detect and correct for small head movements by spatial alignment of all volumes to the first volume by rigid body transformations. Functional slices were co-registered to the anatomical volume using position parameters from the scanner and manual adjustments to obtain optimal fit, and transformed into Talairach space. Modest spatial smoothing was applied to the functional data using a Gaussian filter of 5 mm FWHM.

The functional runs were analyzed using a two-level (hierarchical) multi-subject multiple linear regression (GLM) approach. In the first level analysis, all experimental conditions in all separate subjects were modeled as predictors. Predictor time-courses were adjusted for the hemodynamic response delay by convolution with a double-gamma hemodynamic response function (in which both the BOLD response and the undershoot are modeled by a gamma function; Friston et al., 1998). In the passive experiments (blocked and event-related), the GLM contained four predictors: unimodal visual, unimodal auditory, bimodal congruent and bimodal incongruent. In the active experiment, there were only two conditions: bimodal congruent and bimodal incongruent. Only the correctly answered trials were included in the predictors corresponding to the experimental conditions, incorrectly answered trials were modeled as a separate predictor. This resulted in three predictors: bimodal congruent, bimodal incongruent and incorrect trials.

To explore the effects of presentation mode (blocked vs. random) and task instruction (passive perception vs. active matching), we used a second level (random effects) congruency contrast [congruent vs. incongruent] for all three experiments. This contrast robustly revealed regions in the posterior and anterior auditory association cortex (planum temporale [PT] and anterior superior temporal plane [aSTP]) in our previous fMRI studies. Statistical maps of the congruency contrast were corrected for multiple comparisons using cluster-size thresholding (Forman et al., 1995; Goebel et al., 2006). In this method, an initial voxel-level (uncorrected) threshold is set. Then, thresholded maps are 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 threshold that yields a cluster-level false-positive rate (α) of 5% is applied to the statistical maps. The cluster-size-corrected maps are projected in axial slices of the MNI template brain.

Averaged BOLD-response time-courses of all conditions were extracted in the regions in the auditory association cortex showing overlapping activation for the congruency contrast in the different experiments. To statistically test differences between passive and active integration, we estimated % signal change for congruent and incongruent stimulation for each subject in the two event-related experiments in ROIs selected by the congruency contrast on the passive blocked runs. The mean values were plotted in bar graphs and statistically tested with a 2×2 repeated-measures ANOVA, with experiment (passive, active) and congruency (congruent, incongruent) as within-subject factors.

Study 2: Behavioral experiment

A behavioral experiment was conducted prior to the fMRI experiment to select suitable stimulus onset asynchrony (SOA) values to investigate the effect of task instruction on temporal

windows of integration. Ten healthy native Dutch subjects (2 males, 1 left-handed, mean age 27, range 23–33) participated in the behavioral experiment. All subjects were graduate students, had normal vision and hearing capacity, and no (history of) reading or other language problems.

Task procedure

Subjects were tested individually in a sound-attenuating room, in front of a computer monitor on which the visual stimuli (single letters) were presented. Speech sounds were presented binaurally through loudspeakers. Pairs of letters and speech sounds were presented sequentially (see Experimental trials). After each stimulus pair, subjects had to give a yes/no response to the question: “Are the letter and the sound the same?” by pressing one of two buttons as fast and accurate as possible. The correct button (left/right) for “yes” and “no” was counterbalanced over the subjects.

Experimental trials

In each trial, one letter and one speech sound were presented (Fig. 1A). Congruency (congruent, incongruent) and SOA were randomly distributed over all trials. Eleven SOAs were sampled: –500, –300, –150, –100, –50, 0, 50, 100, 150, 300 and 500 ms. The visual stimulus preceded in trials with a negative SOA, the auditory stimulus preceded in trials with a positive SOA. In total, there were 22 conditions (2 levels of congruency \times 11 levels of SOA), 48 trials were presented per condition, resulting in 1056 trials in total. Stimulus presentation time was 350 ms. Reaction times (RTs) were recorded from the onset of the second stimulus for 1900 ms. The time between the second stimulus of a trial and the start of the next trial was 1900, 2000 or 2100 ms, resulting in a trial duration in the range 1900–2600 ms (see Fig. 1A). Both accuracy (% correct responses) and RT (only correct responses) measures were analyzed using a 2 \times 11 (congruency \times SOA) within-subject ANOVA.

Study 2: fMRI experiment

Participants

Sixteen healthy native Dutch subjects participated in the fMRI experiments in Study 2. Three subjects were excluded from data analyses due to excessive head motion during the functional scans. The remaining 13 subjects (4 males, mean age 23, range 18–34) were all university students enrolled in an undergraduate study program. We selected subjects based on a questionnaire including questions concerning present or past reading or other language problems. All were right-handed, had normal or corrected-to-normal vision, and normal hearing capacity. Subjects gave informed written consent and were paid for their participation.

Experimental procedure

All subjects participated in 2 different experiments (see Table 1): a passive blocked experiment identical to Study 1 (passive experiment), and an active event-related experiment similar to Study 1 but with different temporal offsets (SOAs) between the visual and auditory stimulus (active experiment). The passive experiment was always conducted first, again to ensure that passive perception was uncontaminated with the explicit matching instruction. Based on the results of the behavioral pilot (see Results - Study 2, Fig. 3) and the previous passive fMRI study manipulating temporal asynchrony (Van Atteveldt et al., in press), three SOAs were selected for the present active fMRI experiment: –300 (visual

first, VA), 0 (synchronous, SYN) and +300 (auditory first, AV) ms (see Fig. 1C, Study 2). This resulted in 6 conditions in the active experiment (2 \times 3, congruency \times SOA). Six stimuli of each condition were presented in each run, with a jittered ITI of 13.8, 16.1 or 18.4 s (corresponding to 6, 7, or 8 TRs). Similar to the behavioral experiment and fMRI experiment 1, the subjects were instructed to respond as fast and accurately as possible to the question: “Are the letter and the sound the same?” after each stimulus pair. Reaction times were recorded from the onset of the second stimulus. The correct button (middle or index finger of the right hand) for “yes” and “no” was counterbalanced over the subjects.

Scanning procedure

Imaging was performed on a 3-T head scanner (Magnetom Allegra, Siemens Medical Systems, Erlangen, Germany) located at the M-BIC in Maastricht, The Netherlands. In each subject, 3 or 4 passive blocked runs of 70 volumes were acquired, followed by 3–5 active event-related runs of 262 volumes, using a BOLD-sensitive EPI sequence (scanning parameters: see Study 1). In the passive runs, the scanning sequence was identical to that used in Study 1 (see Fig. 1B). In the active runs, the scanning sequence was also identical to that used in Study 1, except that the effective TR was changed from 2000 ms to 2300 ms, creating a longer delay (800 ms) which was needed for the SOA (300 ms) between the visual and auditory stimulus (Fig. 1C, Study 2). A high-resolution structural scan was collected for each subject using a T1-weighted 3D MP-RAGE sequence identical to Study 1.

Analysis of fMRI time-series

Functional and anatomical images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands; Goebel et al., 2006). The preprocessing procedure was identical to the fMRI experiment in Study 1. For group-level analyses of the functional data of both experiments we used a two-level (hierarchical) multiple linear regression (GLM) approach. In the first level analyses, all experimental conditions in all separate subjects were modeled as predictors (single-factor GLM), or congruency and SOA were modeled as separate factors (two-factor GLM, only in the active experiment). In the second level analyses, random-effects contrasts [congruent vs. incongruent] were performed. Only correctly answered trials were included in the experimental predictors in the active experiment, incorrectly answered trials were modeled as a separate predictor. Predictor time-courses were adjusted for the hemodynamic response delay by convolution with a two-gamma hemodynamic response function. In the passive experiment, the GLM contained four predictors: unimodal visual, unimodal auditory, bimodal congruent and bimodal incongruent. In the active experiment, the single-factor GLM contained seven predictors: bimodal congruent at the three SOAs and bimodal incongruent at the three SOAs, and a predictor for the incorrect trials. In the 2 \times 3 (congruency \times SOA) two-factor GLM, predictors were defined for the main effects of congruency and SOA, the interaction between congruency \times SOA, and for the incorrectly answered trials.

We used two different analytical strategies to infer experimental effects from the fMRI time-series in Study 2. In the first strategy, we performed voxel-wise analyses on the passive runs to select ROIs, and subsequently performed ROI-based analyses on the time-series from the active experiment. These analyses were performed both on a group level and in individual subjects. This first strategy allowed us to observe fMRI responses during active integration at different

SOAs in regions found to be involved in integration during the passive experiment. In the second strategy, we performed voxel-wise analyses directly on the time-series from the active experiment. This second analytical strategy allowed us to explore the effects of SOA and congruency during active integration in regions not identified during passive integration.

First analytical strategy

Voxel-wise multi-subject GLM analyses were performed on the passive runs to replicate our previous findings. Based on the previous results, we searched for regions showing an effect of congruency (congruent > incongruent) in the auditory association cortex. On the group level, random-effects congruency maps were thresholded using the false discovery rate (FDR; Genovese et al., 2002), and one ROI was selected in each hemisphere at $q(\text{FDR}) < 0.05$. In these ROIs, we performed ROI-based analyses on the time-series of the active experiment (i.e., on the time-course of the signal from the whole ROI).

In individual subjects, ROIs were selected using the congruency contrast in single-subject voxel-wise GLM analyses. In each subject, we selected an ROI in each hemisphere based on two criteria: (1) Location, based on Van Atteveldt et al. (2004): posterior auditory association cortex including the planum temporale (PT) and Heschl's sulcus (HS) extending into Heschl's gyrus (HG). Defining individual anatomy was further guided by Duvernoy (1999) and Formisano et al. (2003), see Supplementary Fig. 1. (2) The fMRI response pattern: [Congruent > Auditory > Incongruent > Visual]. Thresholds were adjusted in each subject to yield ROIs with approximately the same cluster size. To examine the experimental effects in the active runs in these individual ROIs, we estimated the % signal change for congruent and incongruent trials at the three SOAs. The mean values were plotted in bar graphs and statistically tested with a 2*3 repeated-measures ANOVA, with congruency (congruent, incongruent) and SOA (VA, Syn, AV) as within-subject factors.

Second analytical strategy

We performed voxel-wise GLM analyses directly on the time-series from the active experiment, to further explore the differences between active and passive integration, and to examine the effect of temporal asynchrony during active letter–sound integration. For these purposes, we tested for voxels showing an interaction between SOA * congruency (using the two-factor GLM), or an effect of congruency (congruent vs. incongruent using the single-factor GLM). Volume-based statistical maps were corrected for multiple comparisons using cluster-size thresholding (Forman et al., 1995; Goebel et al., 2006) and shown in sagittal or transversal slices of the 3D anatomical image of the MNI template brain. Cortex-based statistical maps are shown on the inflated cortical surfaces of the MNI template brain at the same t -value. Clusters showing significant experimental effects were assigned to corresponding anatomical structures using those of Duvernoy (1999).

To summarize, a schematic overview of the different manipulated factors in Studies 1 and 2 is provided in Table 1.

Results

Study 1

Behavioral results

In the active event-related experiment, no significant differences in reaction time were found (congruent: 803.2 ms, incongruent: 829.5 ms, paired-samples t -test: $t = 1.8$, $p = 0.09$). Average accuracy percentages were all above 85%, and also not significantly different (congruent: 91.8%, incongruent 90.7%, $t = 1.02$, $p = 0.33$). Only the fMRI responses to correctly answered trials were analyzed, so the high accuracy levels ensure that only very few trials had to be discarded in the fMRI analysis.

fMRI results

Fig. 2 shows the results of the congruency contrast (congruent vs. incongruent) in the different experiments. Maps are shown at $t(11) = 4$ (p -value at voxel-level = 0.002) with a cluster-size threshold of 148 mm³ (cluster-level corrected at $p < 0.05$). For the passive blocked experiment, extensive regions of superior temporal cortex (STG) bilaterally, including planum temporale (PT) and anterior superior temporal plane (aSTP), were significantly more activated by congruent than by incongruent stimulation (Fig. 2A, orange map, and see Table 2 for details). The congruency contrast also revealed superior temporal regions in the passive event-related experiment (although smaller, see blue map in Fig. 2A), which overlapped considerably with the regions revealed by the blocked experiment. Two separate regions were found in the left hemisphere (PT and aSTP, see Table 2 for details). In the right hemisphere, one posterior region (PT) was revealed; however, this region did not survive the cluster-size threshold procedure. Therefore, this region is not shown in the map, but details are provided in Table 2. In contrast to both passive experiments, no regions in the superior temporal cortex showed an effect of congruency in the active experiment. In Fig. 2A (right), averaged BOLD response time-courses are shown for all three experiments in the regions that overlapped between the blocked and event-related passive experiments. The response patterns are similar for the two passive experiments: [Congruent > Auditory > Incongruent > Visual], while the active experiment showed a different pattern: [Congruent = Incongruent].

The different shape and amplitude of the BOLD response time-courses from the event-related runs are as expected from the stimulus duration differences (see Introduction), as is the smaller cluster size for the passive experiment (see also Buckner et al., 1996). So, the overlapping maps and comparable response patterns for the passive blocked and passive event-related experiments that were found *in spite of* the different stimulation duration and relative amount of scanner noise indicate that an event-related fMRI design is suitable to find the desirable experimental effects.

Fig. 2B shows the fMRI response to congruent and incongruent trials in the passive and active event-related experiments averaged across subjects, in the left and right STG ROIs selected by the congruency contrast in the passive blocked experiment. In both hemispheres, the event-related fMRI response showed a significant

Fig. 2. Results of the congruency contrast in Study 1. (A) *Left*: Random-effects statistical maps of the congruency contrast for both passive experiments (orange: blocked, blue: event-related), projected in transversal slices of the MNI template brain. Maps were thresholded at $t(11) = 4$ (corrected at the cluster-level, $p < 0.05$). No significant activation was found for the same contrast in the active experiment. *Right*: Averaged BOLD response time courses in the regions in which both maps overlapped, shown for the three different experiments. (B) Individually estimated %signal change for congruent and incongruent trials in the passive and active event-related experiments, in the left and right STG ROIs selected by the congruency contrast in the passive blocked experiment (orange activation map in panel A). Error bars indicate S.E.M. across subjects ($n = 12$).

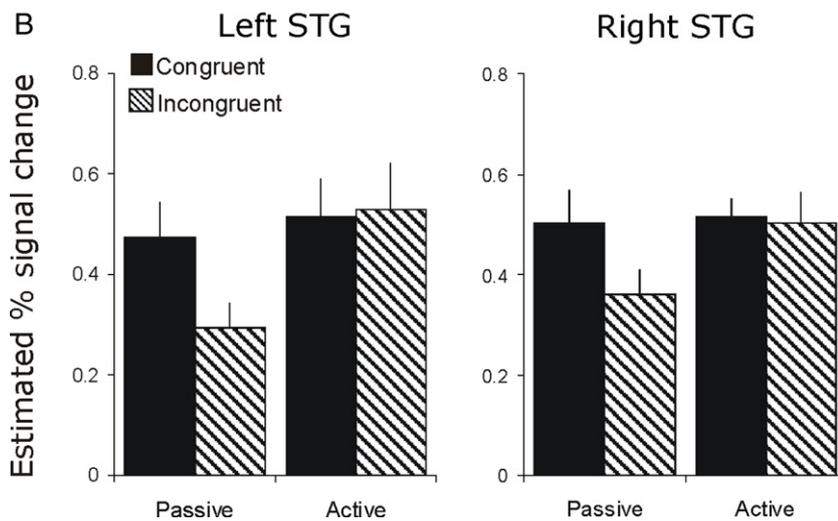
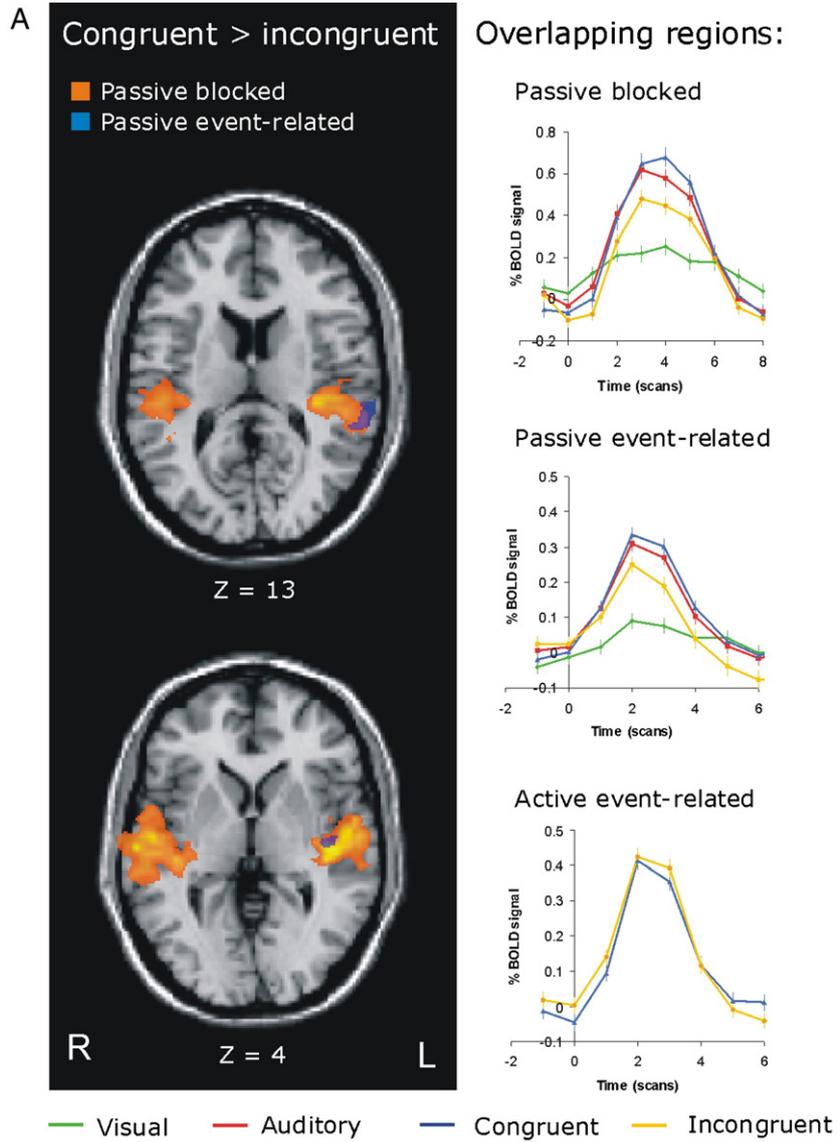


Table 2

Details of the ROIs selected by the [congruent > incongruent] contrast in the different experiments in Study 1

Experiment	ROI		Cluster size ^a	Effect size ^b (voxel level)	
	Anatomy	Talairach (x, y, z)		t	p
Passive blocked	Left STG	(-47, -20, 7)	9828	5.0	0.0007
	Right STG	(49, -22, 5)	12,672	4.8	0.0008
Passive event-related	Left PT	(-59, -33, 12)	447	4.3	0.001
	Left aSTP	(-42, -15, 4)	192	4.1	0.002
	Right PT	(59, -32, 10)	109	4.3	0.001
Active event-related	–	–	–	–	–

^a Cluster-size threshold at $p < 0.05$: 148 mm³.^b *t*- and *p*-values averaged over all voxels in ROI.

interaction between experiment and congruency (Left: $F_{(1,11)} = 13.2$, $p < 0.004$; Right: $F_{(1,11)} = 6.9$, $p < 0.02$) which could be explained by a significant effect of congruency in the passive (Left: $t_{11} = 3.4$, $p < 0.006$, Right: $t_{11} = 3.0$, $p < 0.01$) but not in the active (Left: $t_{11} = -0.5$, $p < 0.6$; Right: $t_{11} = 0.4$, $p < 0.7$) experiment.

Study 2

Behavioral experiment

The 2*11 (congruency*SOA) within-subjects ANOVA with accuracy as dependent variable revealed no significant effects [congruency: $F_{(1,9)} = 0.13$, $p = 0.7$, SOA: $F_{(10,90)} = 1.06$, $p = 0.4$, congruency*SOA: $F_{(10,90)} = 0.88$, $p = 0.5$], all average accuracy levels were above 95%. Using RT as dependent variable, significant experimental effects were observed [congruency: $F_{(1,9)} = 19.2$, $p = 0.002$, SOA: $F_{(10,90)} = 64.3$, $p < 0.000$, congruency*SOA: $F_{(10,90)} = 2.38$, $p = 0.015$]. The effects of SOA and congruency on letter–sound matching reaction times are displayed in Fig. 3.

Fig. 3 shows that reaction times were longest for the synchronously presented letter–sound pairs (SOA=0 ms). In both directions (VA and AV), RTs were reduced with increasing SOA, as shown by the inverted U-shaped curves in Fig. 3 [significant quadratic trend SOA: $F_{(1,9)} = 420.75$, $p < 0.000$]. Reaction times were measured from the onset of the second stimulus, indicating that the decision regarding congruency is faster when information about one of the stimuli is already available. The RT reduction by SOA is stronger when sounds are presented first (AV), as shown by the shorter RTs in the AV direction compared to the VA direction. This difference may be explained by the fact that in the AV direction, the identity information from the second stimulus (visual) is available immediately, while in the VA direction, the second stimulus (auditory) is temporally structured and may therefore take slightly longer to be recognized.

On average, responses to incongruent trials were slower than to congruent trials, however, the congruency*SOA interaction indicates that SOA affected reaction times differently for congruent and incongruent trials. This is also indicated by the significant quadratic trend for congruency*SOA [$F_{(1,9)} = 5.21$, $p = 0.048$]. The RT difference between incongruent and congruent trials seems most pronounced for longer SOAs, especially when letters were presented first. Based on these behavioral results and previous fMRI results (Van Atteveldt et al., 2007), we selected -300 ms (VA) and 300 ms (AV) as SOAs for in the present active fMRI experiment, because the effects of congruency and asynchrony were clearly present at these SOAs.

fMRI experiment

Behavioral results

As in the behavioral experiment, all average accuracy levels were above 95% correct. The high accuracy levels ensured that only very few trials had to be discarded in the fMRI analysis. The effects of SOA and congruency on letter–sound matching reaction times are displayed in Fig. 4. The results of the 2*3 (congruency*SOA) within-subjects ANOVA with RT as dependent variable revealed a significant main effect of SOA but not of congruency [congruency: $F_{(1,12)} = 1.51$, $p = 0.24$, SOA: $F_{(2,24)} = 52.32$, $p < 0.000$], and also no interaction [congruency*SOA: $F_{(2,24)} = 0.36$, $p = 0.70$]. Like in the behavioral experiment, a significant quadratic trend of SOA was found [$F_{(1,12)} = 60.82$, $p < 0.000$], corresponding to the inverted U-shaped curve shown in Fig. 4.

fMRI results: analyses based on passive experiment

Multi-subject analysis. Fig. 5A shows the results of the first analytical strategy on the group level. Consistent with Study 1 and our previous studies, the “congruent > incongruent” contrast on the blocked runs revealed bilateral auditory association areas (pos-

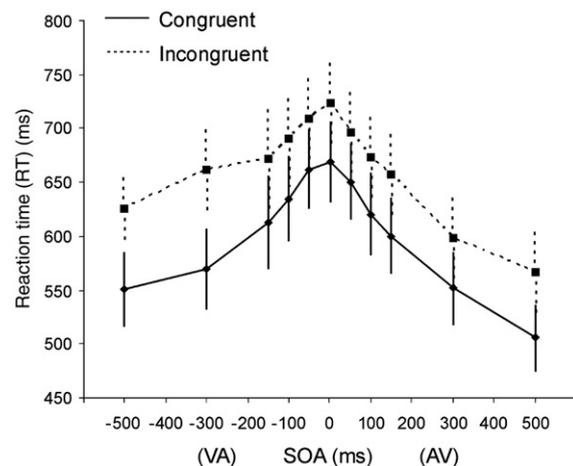


Fig. 3. Reaction time results from the behavioral experiment in Study 2. Average reaction times for congruent (solid line) and incongruent (dashed line) letter–sound pairs are plotted against SOA (stimulus onset asynchrony). Negative SOAs: visual precedes auditory (VA), positive SOAs: auditory precedes visual (AV). Error bars indicate S.E.M. across subjects ($n = 10$).

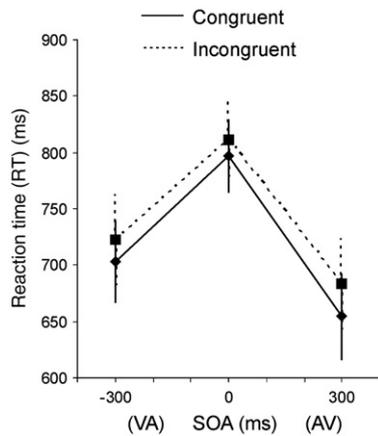


Fig. 4. Reaction time results from the active fMRI experiment in Study 2. Average reaction times for congruent (solid line) and incongruent (dashed line) letter–sound pairs are plotted against SOA (stimulus onset asynchrony). Negative SOAs: visual precedes auditory (VA), positive SOAs: auditory precedes visual (AV). Error bars indicate S.E.M. across subjects ($n=13$).

teriorly on planum temporale anteriorly on aSTP). No significant voxels were detected that showed the opposite contrast (“incongruent>congruent”). One ROI was selected by the congruency contrast in each hemisphere (see Table 3, group ROIs). The averaged time-courses of the BOLD response in the passive experiment showed an auditory specific response (auditory unimodal>visual unimodal), as well as a modulation of the auditory response during bimodal stimulation (congruent>auditory>incongruent). This pattern of enhancement and suppression of the response to speech sounds is a replication of our previous results (Van Atteveldt et al., 2004, 2007) and Study 1. The averaged time-courses of the BOLD response in the active experiment in the same ROIs showed identical responses to congruent and incongruent letter–sound pairs at all SOAs. The results of ROI-based statistics on the time-series of the active experiment confirmed this observation: the difference between congruent and incongruent was not significant at any SOA (Left, VA: $t=-0.997$, $p=0.32$; SYN: $t=-1.696$, $p=0.09$, AV: $t=-1.320$, $p=0.19$; Right, VA: $t=-1.475$, $p=0.14$, SYN: $t=-0.448$, $p=0.65$, AV: $t=-1.455$, $p=0.15$).

Single-subject analysis. Next, we used single-subject GLM analyses to define ROIs based on the congruency contrast in the passive experiment in individual subjects. Because the congruency effect in the posterior auditory association cortex was most robust at the single-subject level (Van Atteveldt et al., 2004), we restricted our individual ROI selection to a predefined posterior region (including planum temporale, Heschl’s sulcus and gyrus, see Materials and methods). The details of the resulting ROIs are given in Table 3 and Supplementary Fig. 1. ROIs located within the predefined anatomical region showing the correct response pattern were found in all 13 subjects in the left hemisphere (shown in Supplementary Fig. 1), and in 11/13 subjects in the right hemisphere. The estimated % signal change in the individual ROIs averaged across subjects is plotted in Fig. 5B for the different conditions and experiments. Fig. 5B shows that also on the single-subject level, no effects of congruency or SOA were found in the active experiment. This observation was confirmed by the within-subjects ANOVA with % signal change as dependent variable

[Congruency: $F_{(1,12)}=0.36$, $p=0.58$ (left), $F_{(1,10)}=0.14$, $p=0.72$ (right); SOA: $F_{(2,24)}=0.81$, $p=0.46$ (left), $F_{(2,20)}=0.29$, $p=0.75$ (right); Congruency * SOA: $F_{(2,24)}=0.006$, $p=0.99$ (left), $F_{(2,20)}=0.01$, $p=0.99$ (right)].

fMRI results: analyses based on active experiment

We found no voxels showing a significant interaction between temporal relation and congruency using the two-factor (SOA * congruency) model. Therefore, we collapsed over SOAs and performed the congruency contrast using the single-factor GLM (see Materials and methods). The resulting statistical map is shown in Fig. 6, at $t(12)=3.2$ (p -value at voxel-level=0.007) with a cluster-size threshold of 407 mm³ (cluster-level corrected at $p<0.05$). As opposed to the passive experiment, only voxels showing the “incongruent>congruent” contrast, but not the “congruent>incongruent” contrast were found in the active experiment. Regions responding stronger to incongruent compared to congruent trials were found in the right parietal cortex and bilateral frontal cortex (Fig. 6A and Table 4). More specifically, frontal activations were found in superior frontal sulcus (SFS), inferior frontal sulcus (IFS) and anterior cingulate gyrus (ACG), parietal activations in superior parietal gyrus (SPG) and the supramarginal gyrus (SMG).

The surface map in Fig. 6A confirms the absence of the congruency effect in the direction “congruent>incongruent” in the auditory association cortex, which was already indicated by the first analytical strategy. In Fig. 6B, averaged time-courses are shown for the parietal and frontal regions exhibiting a stronger response to incongruent trials, pooled over different SOAs (since there was no SOA * congruency interaction), and over hemispheres in IFS. All regions except the SPG showed a response to both congruent and incongruent trials (the response to incongruent trials being stronger), the SPG responded exclusively to the incongruent trials.

Discussion

The principal aim of the present study was to examine the effects of stimulus presentation mode and top–down task-related factors on the neural integration of letters and speech sounds. Based on our previous results (Van Atteveldt et al., 2004, 2007), we explored the effects of content congruency and temporal asynchrony in the auditory association cortex in a series of experiments in which stimulus presentation and task demands were systematically manipulated (see Table 1). The effect of stimulus presentation mode was investigated by comparing blocked and random stimulus presentation in a passive setting (Study 1). The effect of top–down task-related factors was investigated by comparing passive integration to an active version of the same experiment (Study 1) and to an active experiment in which the temporal relation between letters and speech sounds was varied (Study 2).

Effect of stimulus presentation mode

The purpose of the first study was to explore the effect of the stimulus presentation mode on the neural integration of letters and speech sounds. In each subject, three different experiments were conducted in which the experimental design was systematically changed: a passive blocked, a passive event-related, and an active event-related fMRI experiment. In this study, letters and speech sounds were always presented synchronously in the bimodal trials. The direct comparison of the passive blocked and passive event-

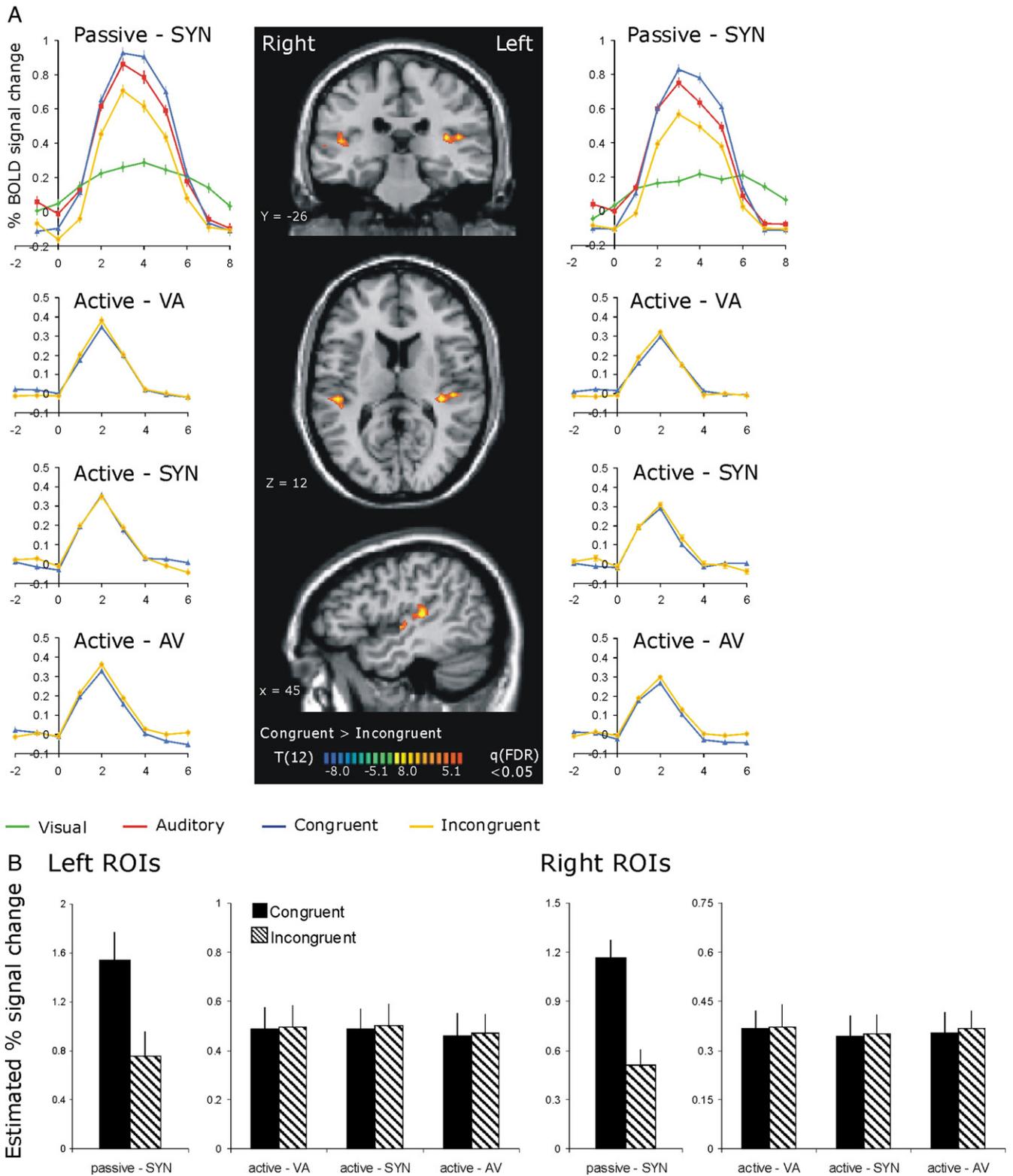


Fig. 5. fMRI results from Study 2: analyses based on the passive experiment. (A) Group results. Middle panel: random-effects statistical map of the congruency contrast, thresholded at $q(\text{FDR}) < 0.05$ and projected on the MNI template brain. Graphs: averaged BOLD response time-courses for the superior temporal regions in left (graphs on the right) and right (graphs on the left) hemisphere. Top row: time courses of the passive experiment (Passive-SYN), rows below: time courses of the active experiment, plotted separately for the different SOAs (Active-VA, Active-SYN, Active-AV). SOA = stimulus onset asynchrony, VA = visual precedes auditory (SOA = -300 ms), SYN = synchronous (SOA = 0 ms), AV = auditory precedes visual (SOA = 300 ms). (B) Single-subject results. Estimated % signal change in the individual ROIs, plotted separately for the passive experiment (Passive-SYN) and the active experiments at different SOAs (Active-VA, Active-SYN, Active-AV). Error bars indicate S.E.M. across subjects ($n = 13$).

Table 3
Details of the group-level and individual ROIs in auditory association cortex defined by [congruent>incongruent] in the passive experiment in Study 2

Subjects	Left			Cluster size	<i>t</i>	<i>p/q</i> ^a	Right			Cluster size	<i>t</i>	<i>p/q</i> ^a
	<i>x</i>	<i>y</i>	<i>z</i>				<i>x</i>	<i>y</i>	<i>z</i>			
<i>Group</i>												
#1–13	−40	−22	8	1015	5.2	0.05	50	−21	6	1477	5.2	0.05
<i>Individual subjects</i>												
#1	−37	−27	16	39	3.0	0.003	49	−25	7	37	2.6	0.01
#2	−39	−21	11	43	3.4	0.0008	51	−21	14	34	3.7	0.0002
#3	−48	−37	16	26	3.2	0.001	54	−27	12	31	3.4	0.0008
#4	−48	−28	20	28	3.2	0.001	53	−16	11	38	3.6	0.0004
#5	−53	−20	7	27	3.2	0.001	–	–	–	–	–	–
#6	−58	−19	17	29	4.2	0.00004	40	−19	9	48	3.0	0.003
#7	−53	−24	6	23	2.3	0.02	46	−13	6	29	2.3	0.02
#8	−63	−23	2	33	4.8	0.000003	47	−21	12	35	3.5	0.0005
#9	−46	−18	4	26	2.8	0.006	51	−25	11	21	2.5	0.01
#10	−34	−28	14	41	3.0	0.003	42	−27	8	42	2.8	0.006
#11	−48	−34	18	23	3.6	0.0003	51	−14	16	36	3.3	0.001
#12	−48	−25	4	23	2.8	0.006	–	–	–	–	–	–
#13	−59	−18	5	33	2.9	0.004	63	−19	6	43	4.2	0.00004
Average	−48.8	−24.8	10.8	30.3			49.7	−20.6	10.2	35.8		
SD	8.6	6.0	6.3	7			6.2	5.0	3.3	7.3		

^a Group ROIs selected at $q(\text{FDR}) < 0.05$, individual ROIs at various p -values (uncorrected).

related experiments addressed two aims: (1) testing the hypothesis that the congruency effect found in the previous studies can be explained by attention or arousal due to the blocked presentation, and (2) controlling for the changes in stimulus presentation mode unrelated to task factors (increased scanner noise and shorter stimulus duration). As shown in Fig. 2, the congruency effect in the auditory association cortex found in previous studies was replicated in both the blocked and the event-related passive designs. Moreover, the anatomical location overlapped and the response patterns were highly similar for the two different stimulus presentation modes. The same order of response strength was shown (Congruent>Auditory>Incongruent>Visual, see Fig. 2B), which was interpreted as a cross-modal modulation of the response to speech sounds depending on the congruency of the letter–sound pairs (Van Atteveldt et al., 2004).

The replication of the congruency effect and the response pattern indicating cross-modal modulation in the “unimodal” auditory association cortex with event-related fMRI has two important implications. First, it indicates that similar to blocks of letter–sound pairs, *single* letter–sound pairs also elicit enhanced and suppressed responses relative to speech sounds alone in the auditory association cortex. This excludes the explanation of the congruency effect in terms of different attention levels in the blocked experiments, because *each* stimulus pair has to be integrated to result in the congruency effect in the event-related experiment.

The second implication is crucial for investigating the effect of top–down task-related factors on the neural integration of letters and speech sounds. The replication of the congruency effect in the auditory association cortex with event-related fMRI excludes the possibility that potentially different response patterns for active integration (using event-related fMRI) and passive integration (using a blocked design) can be attributed to differences in experimental design related to stimulus duration and relative amount of scanner noise. The observation that the same response pattern [Congruent>Auditory>Incongruent>Visual] was present in the event-related passive experiment, *in spite of* differences in scanner

noise and stimulation duration, provides strong support for the suitability of the event-related fMRI experiment to detect differential responses to letters, speech sounds and their congruent and incongruent combinations. Therefore, the intermediate step (passive, event-related) was left out in Study 2, in which the effect of task-related factors was investigated at different temporal offsets between the letter–sound pairs.

Top–down task effects

Fig. 2 shows that in contrast to the passive event-related experiment, the congruency effect was not found in the auditory association cortex during active matching of letters and speech sounds. The identical experimental design used in the passive and active event-related fMRI experiments in Study 1 allows direct comparison of the averaged time-courses, which demonstrates that the absence of the congruency effect seems to be due to an enhanced response to incongruent stimuli during the active experiment. The significance of the different results in both event-related experiments was assessed directly using the individually estimated fMRI responses in the STG–ROIs selected by the blocked experiment (Fig. 2B). We found a significant interaction of experiment and congruency, explained by an effect of congruency in the passive experiment and no effect of congruency in the active experiment. Furthermore, Fig. 2B also shows that the absence of the congruency effect during active matching is due to enhancement of the response to incongruent stimuli.

In Study 2, the effect of top–down task instruction on letter–speech sound integration was investigated at different SOAs. In each subject, two different experiments were conducted: a passive blocked experiment and an active event-related experiment. In the active experiment, letter–sound pairs were presented with different SOAs and the same/different matching task had to be performed. The matching performance during scanning (Fig. 4) resembled the pattern found during a behavioral pilot experiment in which more SOAs were sampled (Fig. 3).

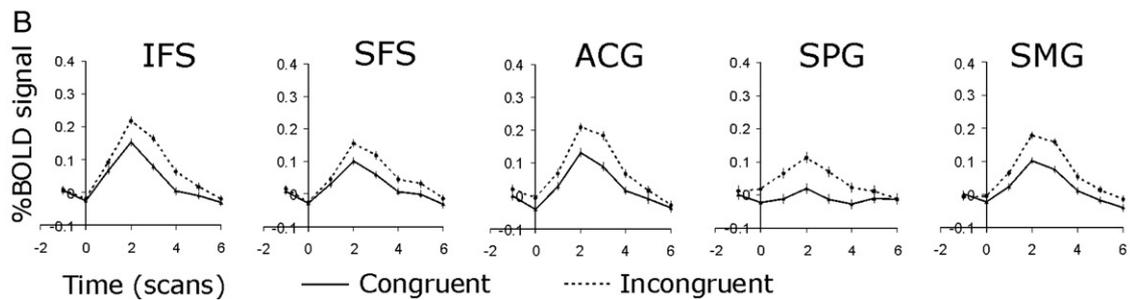
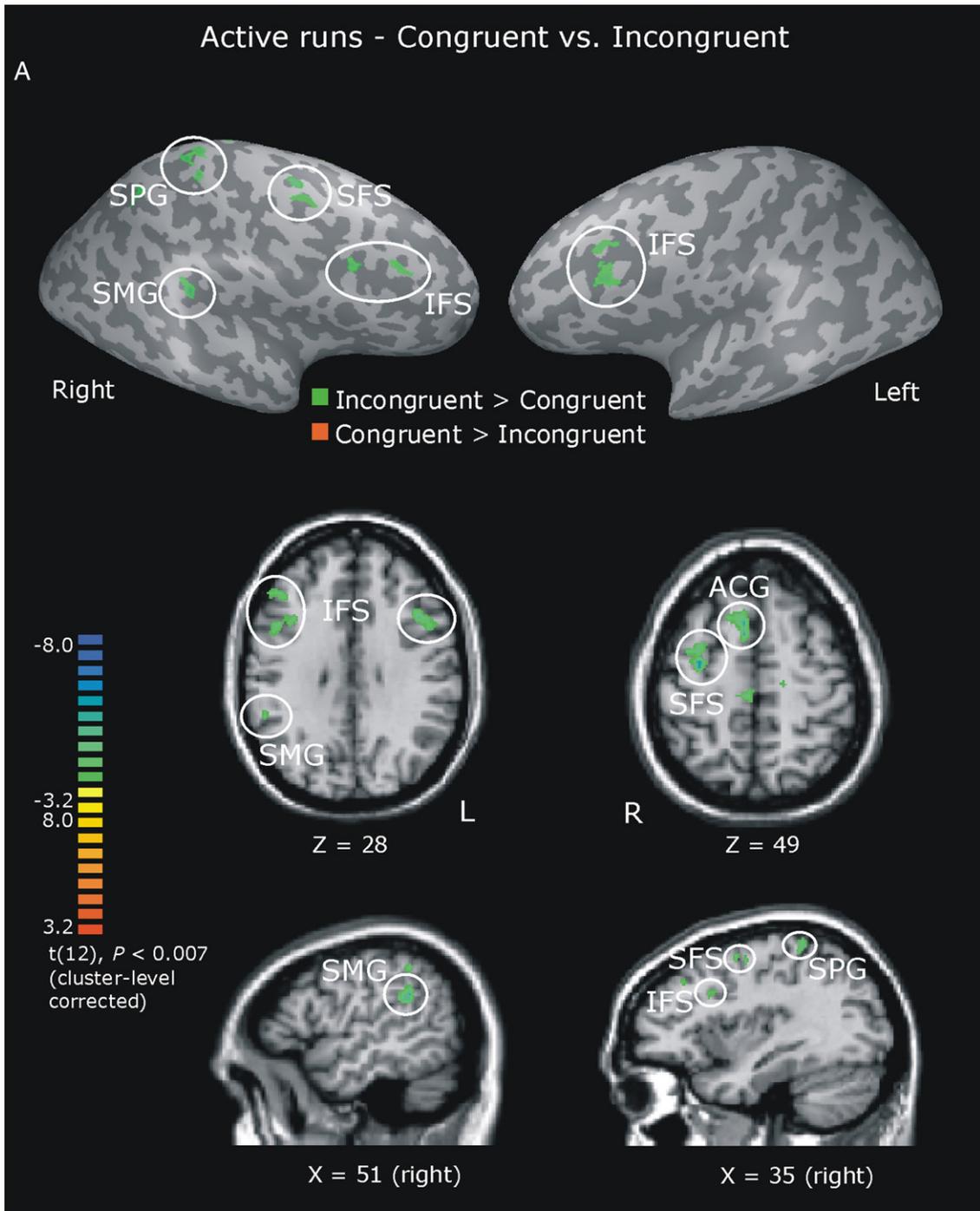


Table 4
Regions revealed by the [incongruent > congruent] contrast in the voxel-wise analysis of the active experiment in Study 2

Anatomical structure	Talairach			Cluster size ^a	Effect size ^b (voxel level)	
	x	y	z		t	p
<i>Frontal</i>						
Right inferior frontal sulcus (IFS)	43	9	27	541	3.6	0.004
Left inferior frontal sulcus (IFS)	−41	11	30	1642	3.6	0.004
Right superior frontal sulcus (SFS)	31	−4	48	617	3.7	0.004
Right anterior cingulate gyrus (ACG)	8	13	48	1169	3.7	0.004
<i>Parietal</i>						
Right superior parietal gyrus (SPG)	23	−31	62	1568	3.7	0.004
Right supramarginal gyrus (SMG)	52	−41	27	508	3.7	0.004

^a Cluster-size threshold at $p < 0.05$: 407 mm³.

^b Average t -value for all voxels in cluster.

We analyzed the functional data from the second study using two different approaches. First, the congruency contrast on the passive runs was used to map ROIs in the auditory association cortex, and ROI-based statistics (congruency contrast at all SOAs) were performed on the active runs in the resulting ROIs. The multi-subject congruency contrast performed on the passive blocked runs replicated the results of our previous studies (Fig. 5A): again, bilateral regions in auditory association cortex (see Table 3 for details) showed the response pattern [Congruent > Auditory > Incongruent > Visual], indicating cross-modal modulation of the response to speech sounds by simultaneously presented letters. In contrast, in the active experiment, the averaged time-courses and ROI-based statistics in the same regions demonstrated identical responses to congruent and incongruent stimuli at all three SOAs. As already indicated in Study 1, no effects of congruency seem to occur in the auditory association cortex when letters and sounds have to be matched explicitly, and apparently, also no effects of SOA.

The congruency contrast on the group level resulted in one large ROI in each hemisphere (Fig. 5A). Because of individual variability in functional neuroanatomy, potential congruency effects in the active experiment on the individual level may have disappeared in the group analysis if they were present in different locations (i.e., in different voxels) within the large ROI. Therefore, the ROI-based group analysis of the active experiment may have lacked sensitivity to detect subtle differences present on the individual level. Therefore, we performed the congruency contrast on the passive runs also in single subjects to define individual ROIs in the auditory association cortex. We restricted our individual ROI selection to a predefined posterior region (including planum temporale, Heschl's sulcus and gyrus, see Materials and

methods), since the posterior auditory association cortex showed the most robust congruency effect at the single-subject level (Van Atteveldt et al., 2004). The averaged Talairach coordinates of the resulting ROIs (Table 3) were remarkably similar to the individual ROIs found for the same contrast in this previous study. Supplementary Fig. 1 shows that the individual ROIs in the left hemisphere are located within the same anatomical region of the posterior auditory association cortex (including PT, HS and HG) in all subjects. Similar results were obtained in the right hemisphere. Fig. 5B shows that also on the individual level, no effects of congruency and SOA were found in the active experiment.

The absence of the effects of congruency and SOA in the active experiment could be due to the restriction of the analyses to regions found by the passive experiment. This possibility was further investigated by the second analytical approach in which voxel-wise analyses were performed on the active runs. No voxels were found that showed an interaction between SOA and congruency, in contrast to our previous passive study (Van Atteveldt et al., 2007). Fig. 6 shows that the congruency contrast (collapsed over SOA) revealed effects only in the direction “incongruent > congruent”, which is consistent with other fMRI results for matching and conflicting audiovisual vowels (Ojanen et al., 2005). Regions responding stronger to incongruent than to congruent letter–sound pairs were located in different regions in the frontal and parietal cortex (Fig. 6A and Table 4), which is also in accordance with the findings of Ojanen et al. (2005), although their reported frontal and parietal activations were more left-lateralized. The whole-brain statistical map shown in Fig. 6A provides further evidence for the absence of the congruency effect in the direction “congruent > incongruent” in the auditory association cortex during active integration of letters and speech sounds.

Passive versus active integration of letters and speech sounds

Taken together, Studies 1 and 2 provide evidence for the major effect of adding an explicit task on the neural mechanism for letter–speech sound integration: the congruency effect in the auditory association cortex, which was the key finding in our previous passive studies, disappeared. Furthermore, the strong effect of temporal asynchrony in the auditory association cortex was also absent during active matching. As demonstrated by the results of Study 1, the absence of the congruency effect seems to be caused by an increased response to incongruent stimuli during the active experiment as compared to the passive experiment (Fig. 2).

The lack of the effects of congruency and SOA suggests that the task instruction changed the default or automatic response pattern in the auditory cortex. In the passive studies on letter–sound integration, subjects perceived the stimuli without any specific task-induced goal. The congruency effect and narrow temporal window of integration in the auditory cortex found during passive perception is therefore likely to result from a default or automatic integration mechanism for letters and speech sounds. In this default perceptual state, the brain uses mainly stimulus-related information (content/identity congruency and temporal coincidence) for deciding which external stimuli belong together and are therefore relevant

Fig. 6. fMRI results from Study 2: voxel-wise analyses of the active experiment. (A) Random-effects statistical maps of the congruency contrast performed on the active runs, superimposed on the inflated cortical surface, and transversal and sagittal slices of the MNI brain. Maps were thresholded at $t(12) = 3.2$ (corrected at the cluster-level $p < 0.05$). (B) Averaged time-courses of the BOLD response to congruent (solid lines) and incongruent (dashed lines) letter–sound pairs in the regions shown in panel A, pooled over SOAs and hemispheres (IFS). SFS: superior frontal sulcus, IFS: inferior frontal sulcus, ACG: anterior cingulate gyrus, SPG: superior parietal gyrus, SMG: supramarginal gyrus.

to process further (Driver and Spence, 2000; Lalanne and Lorenceau, 2004). Congruent and temporally coincident stimuli provide corresponding information about the same external event and will lead to enhanced processing in auditory cortex. On the contrary, incongruent and asynchronously presented combinations of letters and speech sounds are extracted as not belonging together, and will therefore lead to suppressed activity in auditory cortex. As the presumed integration site that determines congruency and temporal relation, the STS/STG may play an important role in signaling this relevance and the subsequent enhancement or suppression of speech sound processing in the auditory cortex.

In the matching task during the active experiments, subjects had to make a same/different decision after each stimulus pair. This task instruction changed the behavioral relevance of the stimuli compared to the passive situation because each stimulus required a decision. Therefore, all stimuli are equally relevant, independent of congruency and temporal offset. Furthermore, in order to perform the task correctly, the congruency information has to be used explicitly in the matching task. The absence of the effects of the stimulus properties (content congruency and SOA) in the auditory cortex during the active task indicates that the default, bottom-up integration process is overruled by the top-down task effects. To summarize, without task instructions, the auditory cortex is modulated to favor the processing of congruent and synchronous information, while this modulation is overruled during an explicit matching task in which congruent and incongruent audiovisual stimuli at different temporal offsets are equally relevant.

A frontal–parietal network as possible source of the top–down task effects

An interesting next question is which network of higher-order brain regions is responsible for the task-related modulation of the default multisensory responses in the auditory association cortex. As shown in Fig. 6, the parietal and frontal regions revealed by the “incongruent vs. congruent” contrast show significant activation in both conditions (except for the SPG), but stronger for incongruent trials. Since mainly the response to *incongruent* letter–sound pairs in the auditory cortex was influenced by the task instruction (see Fig. 2), this frontal–parietal network is a plausible candidate for the source of the top–down task effects. A task-related role for the revealed frontal–parietal network is supported by the observation that these regions did not show a stronger response to incongruent stimuli in the passive experiments. More support is provided by an fMRI investigation of multisensory object integration, in which it was shown that while superior temporal regions were mainly involved in sensory aspects of integration, frontal and parietal activations corresponded more to task-related factors (Beauchamp et al., 2004). Our interpretation of the frontal regions revealed by “incongruent vs. congruent” is unlikely to be confounded by the motor response, since the motor requirements were identical for congruent and incongruent stimuli.

The postulated role for the revealed parietal and frontal regions in task-related top–down modulation of auditory cortex is consistent with previous findings (Hall et al., 2000), and more generally by the involvement of similar regions in goal-directed top–down selection of task-relevant stimuli and responses (Corbetta and Shulman, 2002). The anterior cingulate, prefrontal and inferior frontal cortices have been reported to play an

important role in divided attention, both within (Corbetta et al., 1991) and across modalities (Loose et al., 2003; Johnson and Zatorre, 2006). In the matching task of the present study, both visual and auditory modalities were equally relevant for the task and it was unpredictable in which modality the first stimulus would arrive. The ACG and IFS may therefore be involved in keeping attention divided over the visual and auditory modalities. Alternatively, the IFS cluster may correspond to Broca’s area and may therefore be related to phonetic processing (Ojanen et al., 2005). However, the “incongruent > congruent” response pattern in IFS was exclusively found in the active experiment, which makes a task-related role more plausible. The higher activity during incongruent trials may suggest a role of the ACG in the processing or detection of conflicting information, as reported by other imaging studies (MacDonald et al., 2000; Weissman et al., 2004). However, in contrast to these other studies (which employed Stroop-like tasks), incongruent information in the present same/different matching study does not evoke response conflict.

In another study using speech stimuli, subjects actively integrated written and spoken items at the lexical level (Shaywitz et al., 2001). The activated parietal and inferior frontal regions were suggested to be involved in processing potentially confusable information, which is consistent with the extra recruitment of these regions in the present study during incongruent letters and speech sounds. The involvement of superior and inferior parietal and superior frontal cortices was also reported in a recent fMRI study on cross-modal matching of speech and lip movements (Saito et al., 2005). However, the parietal regions responded stronger to congruent than to incongruent stimuli, which was opposite to our findings and the findings of Ojanen et al. (2005).

General implications: critical role of task instruction

de Gelder and Bertelson (2003) already pointed out that response strategies evoked by experimental task demands may contaminate automatic perceptual integration processes, which complicates the interpretation of behavioral results. Neuroimaging is valuable in this perspective, because integration on the neural level can directly be compared between situations with and without active task demands, while pure behavioral studies of multisensory integration oblige the use of an active task. In the current study, we demonstrated that multisensory effects in the auditory cortex observed during passive integration, were absent during explicit matching of the same audiovisual stimuli. This finding underpins the critical role of top–down task instructions on neuronal response patterns.

The reason why the congruency effect in auditory cortex was completely overruled during the matching task might be that it is a *direct* (or explicit) test of congruency. A critical next experiment would be to use a task in which congruency information is incorporated as an *indirect* factor (de Gelder and Bertelson, 2003). In direct tasks, explicit reflection on the tested aspect may result in (neural) processing different from the automatic perceptual processing. In an indirect task, congruency would not be the factor determining the stimulus-relevance, and might therefore still automatically influence processing in the auditory cortex like in the passive studies. This idea is supported by a recent study from our group in which we used an auditory vowel decision task (Blau et al., 2006). The auditory vowels were combined with congruent or incongruent letters; however, congruency was irrelevant for task performance and therefore an indirect factor. The results show a

strong effect of congruency on reaction times and, critically, also on the fMRI responses in auditory cortex.

Limitations of the present study

Because we did not include within-modality matching conditions (V–V and A–A), the matching task as employed in the present study is not suitable to reveal the brain mechanism specialized for cross-modal matching, like in other studies (Hadjikhani and Roland, 1998; Banati et al., 2000; Saito et al., 2003, 2005; Taylor et al., 2006). However, since the main goal of the present study was to compare response patterns between passive and active integration in prespecified brain regions, within-modality matching conditions were not necessary and would therefore only unnecessarily have increased the duration of the experiments. Another limitation of the present study is that for comparing active and passive integration, it was restricted to bimodal stimulus presentation. A region of high interest, the heteromodal superior temporal cortex (STS/STG), was found in our previous studies to respond stronger to bimodal letter–sound pairs than to unimodally presented letters and speech sounds. Since only bimodal stimuli were presented in the active experiments in the present study, we could not search for regions with such a response profile. To investigate the effect of top–down task-related factors on the integrative role of the STS/STG, other experiments will be necessary using a different behavioral task, e.g. including within-modality matching conditions. Finally, the comparison of passive and active integration might have been more direct if an event-related fMRI design had been used in the passive experiment in Study 2. However, for achieving equal statistical power, the duration of slow event-related runs are much longer than that of blocked runs. As our previous studies showed the robustness and reproducibility of the congruency effect using a passive blocked design, and the congruency effect was replicated using a passive event-related design in the first study, we chose to use the passive blocked experiment in Study 2 to confidently map regions involved in passive integration.

Conclusions

In the first study, we replicated with event-related fMRI the congruency effect in the auditory association cortex found previously with blocked fMRI. The demonstration of the congruency effect in the auditory association cortex with random stimulus presentation refutes an explanation by attention or arousal due to the blocked stimulus presentation in these previous studies. Furthermore, this finding shows that event-related fMRI is suitable for studying integration of letters and speech sounds, since differential responses to single letters, speech sounds and their (congruent or incongruent) combinations can be detected.

The present studies furthermore showed that the effects of congruency (Studies 1 and 2) and temporal asynchrony (Study 2) in the auditory association cortex were absent during active integration. This indicated that top–down task instructions overruled the effects of automatic, bottom–up integration of letters and speech sounds. Because of the task instruction, all stimulus pairs were equally relevant in the active experiment, which resulted in the disappearance of the effects of congruency and SOA in the auditory cortex. The source of this top–down task effect was suggested to be a frontal–parietal network of attention-related brain areas, controlling the explicit cross-modal integration of the over-learned associations between letters and speech sounds.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.03.065.

References

- Amedi, A., von Kriegstein, K., Van Atteveldt, N.M., Beauchamp, M.S., Naumer, M.J., 2005. Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res.* 166, 559–571.
- Andersen, T.S., Tiippana, K., Sams, M., 2004. Factors influencing audiovisual fission and fusion illusions. *Cogn. Brain Res.* 21, 301–308.
- Banati, R.B., Goerres, G.W., Tjoa, C., Aggleton, J.P., Grasby, P., 2000. The functional anatomy of visual–tactile integration in man: a study using positron emission tomography. *Neuropsychologia* 38, 115–124.
- Beauchamp, M., 2005. See me, hear me, touch me: multisensory integration in lateral occipital–temporal cortex. *Curr. Opin. Neurobiol.* 15, 1–9.
- Beauchamp, M., Lee, K., Argall, B., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823.
- Blau, V., van Atteveldt, N.M., Formisano, E., Goebel, R., Blomert, L., 2006. Visual influences on speech sound discrimination: a parametric fMRI study. *NeuroImage* 31, S38.
- Buckner, R.L., Bandettini, P.A., O’Craven, K.M., Savoy, R.L., Petersen, S.E., Raichle, M.E., Rosen, B.R., 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci.* 93, 14878–14883.
- Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123.
- Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev., Neurosci.* 3, 201–215.
- Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E., 1991. Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *J. Neurosci.* 11, 2383–2402.
- de Gelder, B., Bertelson, P., 2003. Multisensory integration, perception and ecological validity. *Trends Cogn. Sci.* 7, 460–467.
- Driver, J., Spence, C., 2000. Multisensory perception: beyond modularity and convergence. *Curr. Biol.* 10, R731–R735.
- Duvernoy, H.M., 1999. *The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply*. Springer-Verlag, Wien New York.
- Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W., Mintun, M., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.
- Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., Goebel, R., 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40, 859–869.
- Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., Turner, R., 1998. Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40.
- Genovese, C., Lazar, N., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878.
- Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with BrainVoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–402.
- Hadjikhani, N., Roland, P.E., 1998. Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study. *J. Neurosci.* 18, 1072–1084.

- Hall, D.A., Haggard, M.P., Akeroyd, M.A., Summerfield, A.Q., Palmer, A.R., Elliott, M.R., Bowtell, R.W., 2000. Modulation and task effects in auditory processing measured using fMRI. *Hum. Brain Mapp.* 10, 107–119.
- Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. *NeuroImage* 15, 733–746.
- Johnson, J.A., Zatorre, R.J., 2006. Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage* 31, 1673–1681.
- King, A.J., Calvert, G.A., 2001. Multisensory integration: perceptual grouping by eye and ear. *Curr. Biol.* 11, R322–R325.
- Lalanne, C., Lorenceau, J., 2004. Crossmodal integration for perception and action. *J. Physiol. (Paris)* 98, 265–279.
- Loose, R., Kaufmann, C., Auer, D.P., Lange, K.W., 2003. Human prefrontal and sensory cortical activity during divided attention tasks. *Hum. Brain Mapp.* 18, 249–259.
- MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S., 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838.
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I.P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *NeuroImage* 25, 333–338.
- Saito, D.N., Okada, T., Morita, Y., Yonekura, Y., Sadato, N., 2003. Tactile–visual cross-modal shape matching: a functional MRI study. *Cogn. Brain Res.* 17, 14–25.
- Saito, D., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., Sadato, N., 2005. Cross-modal binding and activated attentional networks during audiovisual speech integration: a functional MRI study. *Cereb. Cortex* 5, 1750–1760.
- Shaywitz, B.A., Shaywitz, S.E., Pugh, K.R., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable, R.T., Marchione, K.E., Fletcher, J.M., Klorman, R., Lacadie, C., Gore, J.C., 2001. The functional neural architecture of components of attention in language-processing tasks. *NeuroImage* 13, 601–612.
- Taylor, K.I., Moss, H.E., Stamatakis, E.A., Tyler, L.K., 2006. Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci.* 102, 8239–8244.
- Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282.
- Van Atteveldt, N.M., 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.
- Weissman, D.H., Warner, L.M., Woldorff, M.G., 2004. The neural mechanisms for minimizing cross-modal distraction. *J. Neurosci.* 24, 10941–10949.