

The effect of temporal asynchrony on the multisensory integration of letters and speech sounds

Citation for published version (APA):

van Atteveldt, N. M., Formisano, E., Blomert, L. P. M., & Goebel, R. W. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, 17(4), 962-974. <https://doi.org/10.1093/cercor/bhl007>

Document status and date:

Published: 01/01/2007

DOI:

[10.1093/cercor/bhl007](https://doi.org/10.1093/cercor/bhl007)

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.

The Effect of Temporal Asynchrony on the Multisensory Integration of Letters and Speech Sounds

Nienke M. van Atteveldt¹, Elia Formisano¹, Leo Blomert¹ and Rainer Goebel^{1,2}

¹Department of Cognitive Neuroscience, University of Maastricht, 6200 MD Maastricht, The Netherlands and

²F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, 6500 HB Nijmegen, The Netherlands

Temporal proximity is a critical determinant for cross-modal integration by multisensory neurons. Information content may serve as an additional binding factor for more complex or less natural multisensory information. Letters and speech sounds, which form the basis of literacy acquisition, are not naturally related but associated through explicit learning. We investigated the relative importance of temporal proximity and information content on the integration of letters and speech sounds by manipulating both factors within the same functional magnetic resonance imaging (fMRI) design. The results reveal significant interactions between temporal proximity and content congruency in anterior and posterior auditory association cortex, indicating that temporal synchrony is critical for the integration of letters and speech sounds. The temporal profiles for multisensory integration in the auditory association cortex resemble those demonstrated for single multisensory neurons in different brain structures and animal species. This similarity suggests that basic neural integration rules apply to the binding of multisensory information that is not naturally related but overlearned during literacy acquisition. Furthermore, the present study shows the suitability of fMRI to study temporal aspects of multisensory neural processing.

Keywords: audiovisual, auditory cortex, fMRI, STS, temporal proximity

Introduction

In the natural environment, multisensory stimuli arising from the same event are in close temporal proximity. Not surprisingly, temporal correspondence is a key determinant for the binding of information from different modalities, as is demonstrated in multisensory neurons in the superior colliculus and cortex in the cat (Meredith and others 1987; Stein and Wallace 1996) as well as in primates (Wallace and others 1996). In accordance to these neurophysiological data, the importance of temporal coincidence has also recently been demonstrated using high-resolution functional magnetic resonance imaging (fMRI) in the monkey auditory cortex (Kayser and others 2005).

In these studies, simple transient stimuli such as light flashes and noise bursts are commonly used (for a review, see Stein and Meredith 1993). When the complexity of multisensory information increases, information content of the unisensory inputs may serve as an additional binding factor (Calvert and others 1998; Pourtois and de Gelder 2002; Laurienti and others 2004). Multisensory information may even be exclusively related by information content, for example, when the unisensory inputs are not naturally related. Studies using complex natural multisensory materials that share information content in addition to temporal onset, such as audiovisual speech, have shown that a larger temporal disparity is allowed before integration is disrupted (Massaro and Cohen 1993; Massaro and others 1996;

Munhall and others 1996; Munhall and Vatikiotis-Bateson 2004). Taken together, the importance of temporal proximity seems to depend on the nature and complexity of the multisensory information. We investigated the role of temporal proximity on the integration of letters and speech sounds, which are not naturally related but explicitly learned during literacy acquisition and therefore initially only related by information content.

In speech-based alphabetic scripts, letters and speech sounds are the basic elements of correspondence between written and spoken language. Therefore, learning the correspondences between the letters and speech sounds of a language is a crucial step in literacy acquisition (Ehri 2005). In literate adults, letter-speech sound associations can be considered as overlearned paired associates. However, developmental dyslexics encounter problems learning the correspondences between letters and speech sounds, which is thought to be one of the main causes underlying their reading difficulties (Vellutino and others 2004). Taken together, it is of great relevance to elucidate the role of temporal proximity in the neural binding of letters and speech sounds, both for a better understanding of the principles underlying multisensory integration in the human brain as well as considering the important role of letter-sound correspondences in alphabetic literacy.

In a previous fMRI study, we demonstrated that heteromodal superior temporal regions (superior temporal gyrus [STG] and superior temporal sulcus [STS]) and modality-specific posterior auditory association cortex (planum temporale [PT]) are crucially involved in the neural binding of letters and speech sounds (Van Atteveldt and others 2004). In the present study, we used fMRI to address the question how these multisensory effects in the auditory association cortex and heteromodal STS/STG are influenced by a temporal offset between the letters and speech sounds. For this purpose, we manipulated both the temporal relation (stimulus onset asynchrony [SOA]) and content congruency (same/different identity) between letters and speech sounds within the same experimental design.

As substantiated in recent methodological and review papers, multisensory fMRI results should be interpreted with caution (Calvert 2001), especially when the criterion of superadditivity is used (Beauchamp 2005b; Laurienti and others 2005). One of the main reasons for this is that with fMRI, large amounts of neurons are sampled simultaneously, which complicates the inference of integrative operations on the neuronal level and thereby the use of criteria derived from electrophysiological studies. Another important reason is that because of the intrinsic nature of the blood oxygenation level-dependent (BOLD) response and its limited dynamic range, a superadditive response at the neuronal level is not necessarily reflected in a superadditive change of the BOLD fMRI signal.

We used a congruency effect (at different SOAs) to determine the influence of temporal relation on multisensory integration. In this analysis, 2 bimodal conditions are contrasted to each other, one in which the stimuli have the same identity (congruent) and one in which the stimuli are of different identity (incongruent). The congruency effect can be used as a criterion for multisensory integration because a distinction between corresponding and noncorresponding letters and speech sounds cannot be established unless the unisensory inputs have been integrated successfully. An important advantage of using the congruency effect is that it allows manipulation of the temporal relation between the bimodal stimuli within the same design. Interactions between temporal relation and congruency therefore directly demonstrate an influence of temporal relation on multisensory integration.

Regions exhibiting a congruency effect are not necessarily performing integrative operations themselves, as it cannot be excluded that this effect may reflect feedback from a different region where integration takes place (Van Atteveldt and others 2004). To gain more detailed insight in the functional properties of different regions involved in letter-speech sound integration, it is important to inspect unimodal responses in candidate integration regions (Wright and others 2003; Beauchamp 2005b). Therefore, we presented letters and speech sounds also unimodally. This enabled additional analyses using the criterion that bimodal responses should exceed both unimodal responses (Van Atteveldt and others 2004). This criterion was termed the “max criterion” by Beauchamp (2005b).

In analogy to electrophysiological studies (Meredith and Stein 1983; Meredith and others 1987; Stein and Wallace 1996; Wallace and others 1996), we visualized the magnitude of multisensory interaction (MSI) at different SOAs in regions of interest (ROIs) revealed by the Congruency \times SOA interaction and the max criterion. In electrophysiology, MSI has been defined as a significant difference between the number of impulses evoked by a multisensory stimulus and the number of impulses evoked by the most effective unisensory stimulus, which can either be an enhancement or depression (Stein and others 2004). Although the nature of the measured signal in the present study is evidently different, the same definition is conceptually attractive to quantify and visualize the effect of SOA on multisensory fMRI responses.

Materials and Methods

Participants

Eight healthy native Dutch subjects (7 female, mean age 23 years, range 19–29 years) participated in the present study. All subjects were university students enrolled in an undergraduate study program. Subjects without history of reading or other language problems were selected on the basis of a questionnaire. All subjects 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.

Stimulation Procedure

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; vowels and consonants were presented in separate blocks). 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) rather than 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 (± 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 fixation periods and scanning, a white fixation cross was presented in the center of the screen.

A schematic description of the experimental design is shown in Figure 1. Letters and speech sounds were presented in blocks of unimodal or bimodal stimulation. Congruency (congruent vs. incongruent) and temporal relationship (SOA) between the letters and speech sounds were systematically varied over the bimodal stimulation blocks. Five different SOAs were sampled: -300 , -150 , 0 , 150 , and 300 ms (onset of the letter relative to onset of the sound). In total, there were 12 experimental conditions: unimodal visual, unimodal auditory, bimodal congruent at 5 SOAs, and bimodal incongruent at 5 SOAs. Subjects passively listened to and/or viewed the stimuli to avoid interaction between activity related to stimulus processing and task-related activity due to cognitive factors.

To avoid interference of scanner noise with experimental auditory stimulation, stimuli were presented in silent delay periods between subsequent whole-brain scans (see Fig. 1). Experimental blocks (24 s) were composed of 4 miniblocks of 6 s each. One whole-brain scan was acquired in the beginning of each miniblock, during which only a fixation cross was presented. In the subsequent silent delay, 5 stimuli were presented with an intertrial interval of 800 ms. Because stimulus perception is uncontaminated by scanner noise in the silent period between successive scans, this stimulation procedure is very suitable for studying auditory processing with fMRI (Jäncke and others 2002; Van Atteveldt and others 2004). Stimulus presentation was synchronized with the scanner pulses using the software package “Presentation” (<http://neurobehavioral.com>). Four repetitions of each of the 12 conditions were distributed over 4 experimental runs, resulting in the presentation of 80 trials per condition. The order of the conditions was randomized within runs and counterbalanced across runs. Fixation periods were presented in the beginning and end of each run (36 s) and between each experimental block (24 s).

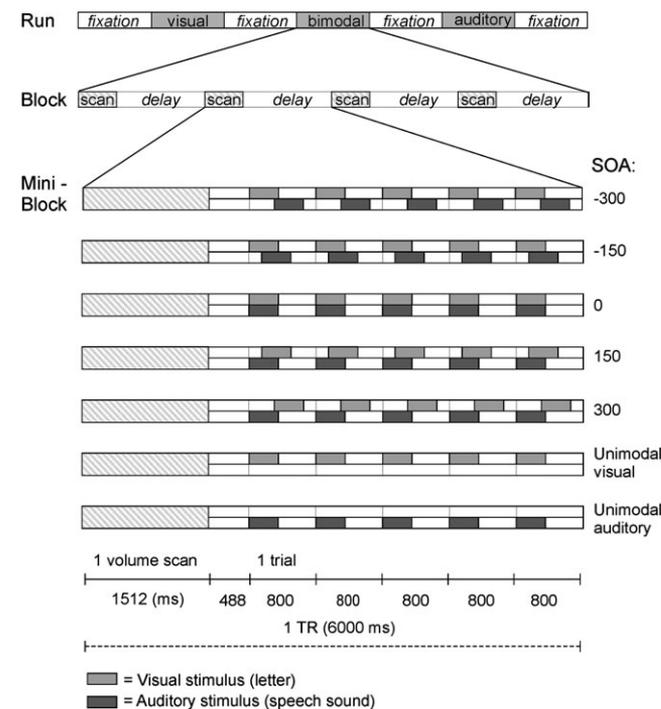


Figure 1. Schematic description of the experimental design. Experimental blocks of 24 s consisted of 4 miniblocks of 6 s. Each miniblock started with the acquisition of one whole-brain scan (1512 ms) followed by 5 experimental trials (ITI = 800 ms) in a silent delay before the next scan was acquired. In the bimodal blocks, each trial consisted of a visual and an auditory stimulus, which were presented with 5 different SOAs (one SOA per block). The timing details within one miniblock are depicted separately for each block type. ITI, intertrial interval.

Scanning Procedure

Imaging was performed on a 3-T whole-body system (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany). In each subject, 4 runs of 104 volumes were acquired using a BOLD-sensitive echo planar imaging sequence (matrix: $64 \times 64 \times 24$, voxel size: $3.5 \times 3.5 \times 4.5 \text{ mm}^3$, field of view: 224 mm^2 , echo time [TE]/repetition time [TR] slice = $32/63 \text{ ms}$, flip angle [FA] = 75°). Sequence scanning time was 1512 ms, and interscan gap was 4488 ms, resulting in a TR (sequence repeat time) of 6000 ms. A slab of 24 axial slices (slab thickness: 10.8 cm) was positioned in each individual such that the whole brain was covered, based on anatomical information from a scout image of 7 sagittally oriented slices. A high-resolution structural scan (voxel size: $1 \times 1 \times 1 \text{ mm}^3$) was collected for each subject using a T_1 -weighted 3-dimensional (3D) magnetization prepared rapid acquisition gradient echo (MP-RAGE) sequence (TR = 2.3 s, TE = 3.93 ms, 192 sagittal slices).

Analysis of fMRI Time Series

Functional and anatomical images were analyzed using BrainVoyager 2000 and BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). 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. Estimated translation and rotation parameters were inspected and never exceeded 1 mm. Functional slices were coregistered to the anatomical volume using position parameters from the scanner and manual adjustments to obtain optimal fit and transformed into Talairach space. No spatial smoothing was applied to the fMRI data.

For visualization of the statistical maps, all individual brains were segmented at the gray/white matter boundary (using a semiautomatic procedure based on intensity values), and the cortical surfaces were reconstructed and inflated. To improve the spatial correspondence mapping between subjects' brains beyond Talairach space matching, the reconstructed cortices were aligned using curvature information reflecting the gyral/sulcal folding pattern (cortex-based alignment procedure, described in Van Attevelde and others 2004). Statistical maps shown in slices are all thresholded using the false discovery rate (FDR) at $q < 0.05$ (Genovese and others 2002).

The fMRI time series were analyzed using 2 differently specified multisubject fixed-effects general linear models (GLMs). In the first GLM, all 12 conditions were modeled as separate predictors (GLM1). The second was a 2×5 factorial model with the factors Congruency (congruent, incongruent) and SOA ($-300, -150, 0, 150, 300 \text{ ms}$), including the interaction term (Congruency \times SOA) and separate predictors for the 2 unimodal conditions (GLM2). Predictor time courses were adjusted for the hemodynamic response delay by convolution with a hemodynamic response function (Boynton and others 1996).

We used GLM1 to contrast all conditions against baseline to create statistical maps of the areas activated by letters, speech sounds, and their combined presentation (Fig. 2). Furthermore, we performed the contrasts (bimodal congruent > bimodal incongruent) at all 5 SOAs using GLM1 (referred to as "congruency contrast" in Results). Clusters for which the congruency contrast was significant (at $q[\text{FDR}] < 0.05$) were saved as ROIs (specified in Table 1). A third analysis performed with GLM1 is the conjunction of [(bimodal congruent > unimodal auditory) \cap (bimodal congruent > unimodal visual) \cap (unimodal auditory > baseline) \cap (unimodal visual > baseline)] (referred to as "max criterion analysis" in Results). In this conjunction analysis, a new statistical value was computed for each voxel as the minimum of the statistical values obtained from the 4 included contrasts (Van Attevelde and others 2004). Clusters for which this new statistical value was significant (at $q[\text{FDR}] < 0.05$) were saved as ROIs. GLM2 was used to reveal interactions between Congruency and SOA (referred to as "interaction analysis" in Results). Clusters that showed a significant interaction (at $q[\text{FDR}] < 0.05$) between Congruency and SOA were saved as ROIs. In addition, we performed the same GLM1 and GLM2 analyses in individual subjects. Individual ROIs were selected at a more liberal threshold ($P < 0.05$).

In the ROIs selected on basis of the multisubject analyses, we estimated individual magnetic resonance (MR) signal levels during the experimental conditions as percentage of the average MR level during fixation periods (baseline). We used these percent signal values to visualize the response pattern at SOA = 0 to provide additional information about intersubject variability of the experimental effects. Furthermore, we used the estimated MR signal levels to calculate MSI values to quantify multisensory integration effects. The magnitude of MSI is calculated by the formula: $((AV - [A, V]_{\text{max}}) / [A, V]_{\text{max}}) \times 100\%$, where AV is the bimodal response and $[A, V]_{\text{max}}$ the most effective unimodal response (Meredith and Stein 1983; Meredith and others 1987; Stein and Wallace 1996; Wallace and others 1996). We used the total percent signal values (baseline [100%] + signal change, e.g., 101.4%), for the calculations of the MSI instead of the percent signal change (e.g., 1.4%), to avoid the MSI to reach extremely high values for occasionally very low (approaching 0) maximal unimodal responses.

Results

Overview of Activated Brain Regions

Figure 2 shows an overview of activated brain regions during the different unimodal (Fig. 2A) and bimodal (Fig. 2B) stimulation periods after cortex-based alignment of anatomical and functional data (see Materials and Methods). In the bimodal conditions, 5 different SOAs were used (SOA between the letter and speech sound). Negative SOAs indicate that the letter was presented first (VA), positive SOAs that the sounds were presented first (AV). At SOA = 0, letters and speech sounds were presented in synchrony (synch).

Figure 2 shows that letters and speech sounds activated similar occipital and temporal brain regions in all different conditions used in the present study. Furthermore, the occipital and temporal activations were consistent with our previous study (Van Attevelde and others 2004) and with other findings: single letters activated extrastriate lateral occipital cortex (e.g., Longcamp and others 2003; Flowers and others 2004), and speech sounds activated anterior as well as posterior superior temporal cortex (see Arnott and others 2004; Scott 2005). Interestingly, the maps for unimodally presented letters and speech sounds overlapped in the STS (Fig. 2A, intersection auditory \cap visual), indicating multisensory convergence of letter and speech sound processing in this region.

In addition to occipital and temporal activations, activated areas were also observed in pre- and postcentral gyri and inferior parietal cortex, with comparable patterns for all unimodal and the bimodal conditions. The activation of the precentral gyrus was most prominent and consistent across conditions. Activation of premotor areas by passive listening to speech sounds is consistent with other findings (Wilson and others 2004) and suggests an influence of articulatory features on speech perception. The premotor regions activated by passive viewing of single letters may correspond to Exner's area, which is thought to be the motor center of writing (Longcamp and others 2003; Matsuo and others 2003).

Congruency Contrast

For synchronous presentation, activation of superior temporal cortex by congruent stimulation was increased compared with incongruent stimulation (see Fig. 3). Interestingly, this difference was absent or less pronounced for the asynchronous conditions: only the contrast map at SOA = 0 revealed significant differences between congruent and incongruent stimulation in

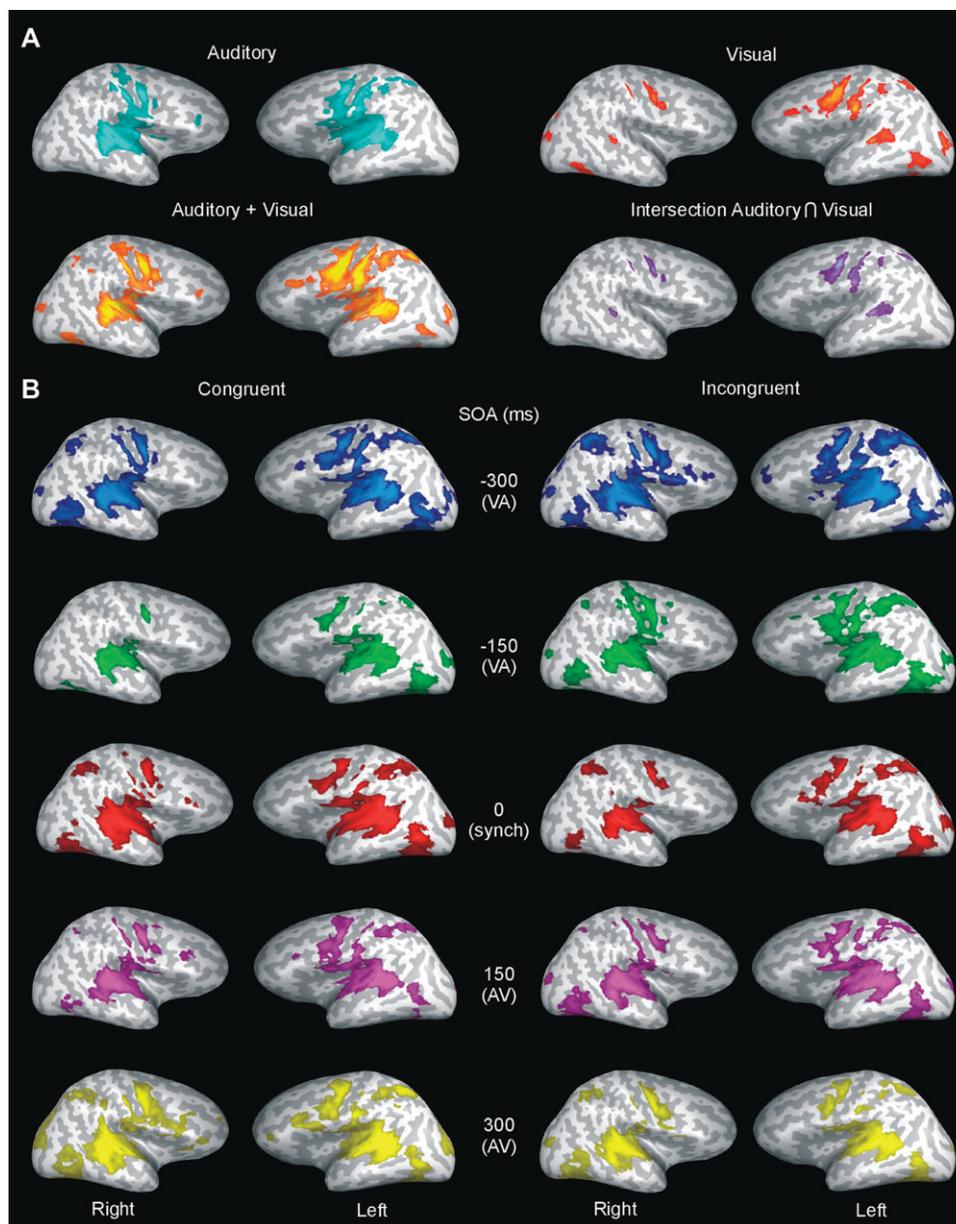


Figure 2. Overview of activation patterns for unimodal and bimodal presentation of letters and speech sounds. (A) Multisubject GLM1 maps of the unimodal predictors against baseline (upper row, left: [auditory > baseline], right: [visual > baseline]), the combined unimodal predictors (lower row, left: [(auditory + visual) > baseline]), and the intersection of both unimodal predictors against baseline (lower row, right: [visual > baseline] ∩ [auditory > baseline]). (B) Multisubject GLM1 maps of the bimodal predictors against baseline for the 5 different SOAs: -300 (blue map), -150 (green map), 0 (red map), 150 (violet map), and 300 (yellow map). Negative SOAs indicate that the letter was presented first (VA), and positive SOAs indicate that the sounds were presented first (AV). At SOA = 0, letters and speech sounds were presented in synchrony (synch). The maps were created from cortex-based aligned functional data and shown on the inflated cortical sheet of the individual brain used as target for the alignment.

the superior temporal cortex (Fig. 3A, orange activation map). The location (Table 1 and Fig. 3A,B) and response patterns (Fig. 3C) of the posterior regions correspond to those observed in the PT in our previous study. In addition, we found a similar response pattern in anterior auditory association cortex bilaterally (anterior superior temporal plane [aSTP], Fig. 3A,B). Individual analyses (congruency contrast at SOA = 0 using GLM1) revealed PT regions in 7/8 subjects in the left hemisphere (average Talairach coordinates \pm standard error of mean [SEM]: -58 ± 2 , -29 ± 4 , 15 ± 1) and in 7/8 subjects in the right hemisphere (61 ± 1 , -24 ± 3 , 15 ± 2); aSTP regions in 8/8 subjects in the left hemisphere (average Talairach coordinates \pm

SEM: -56 ± 2 , -8 ± 1 , 6 ± 1) and in 7/8 subjects in the right hemisphere (58 ± 2 , -8 ± 2 , 3 ± 2).

The averaged BOLD response time courses in Figure 3C indicate that in the PT as well as in the aSTP, the response to congruent letter-sound pairs was stronger than to speech sounds presented in isolation, whereas the response to incongruent letter-sound pairs was weaker than to isolated speech sounds. This observation was confirmed by ROI-GLM analyses for congruent > auditory in right PT ($P < 0.005$), left aSTP ($P < 0.005$) and right aSTP ($P < 0.01$), and marginally in left PT ($P < 0.1$). ROI-GLM results of the auditory > incongruent contrast was only significant in left PT and right aSTP ($P < 0.05$),

Table 1
Details of the ROIs selected by the different analyses

ROI	Hemisphere	Talairach	Number of voxels	Effect size ^a		Statistical test ^b
				<i>T</i>	<i>P</i>	
PT	Left	-52, -31, 15	36	3.42	0.0007	Congruent > incongruent ^c
	Right	60, -20, 16	167	3.74	0.0004	
aSTP	Left	-55, -8, 7	127	3.61	0.0005	Interaction SOA × Congruency
	Right	64, -9, 6	139	4.04	0.0002	
PT	Left	-54, -31, 16	13	3.35	0.0008	
	Right	60, -20, 15	101	3.59	0.0004	
aSTP	Left	-56, -7, 7	189	3.72	0.0003	
	Right	63, -10, 6	194	4.22	0.0002	
STS	Left	-55, -43, 13	11	3.47	0.0005	(Bimodal > unimodal) ∩ (unimodal > 0) ^c

^aEffect size = average *t*-value and *P* value for all voxels in the ROI.

^bStatistical test used for ROI selection (maps thresholded at $q[\text{FDR}] < 0.05$).

^cThese tests were performed at SOA = 0.

approaching significance in the left aSTP ($P < 0.1$), and not significant in the right PT ($P = 0.2$).

Interaction Analysis

Analysis of the fMRI time series using a 2×5 factorial model (GLM2, see Materials and Methods) revealed significant interactions between Congruency and SOA in posterior (PT) and anterior (aSTP) auditory association cortex bilaterally (Fig. 4A). These regions were identical to those revealed by the congruency contrast at SOA = 0 (see Table 1). Individual analyses using GLM2 revealed a significant Congruency × SOA interaction in PT in 8/8 subjects in the left hemisphere (average Talairach coordinates ± SEM: -56 ± 3 , -31 ± 4 , 14 ± 1) and in 6/8 subjects in the right hemisphere (61 ± 1 , -25 ± 1 , 15 ± 2). In aSTP, individual analyses revealed a Congruency × SOA interaction in 7/8 subjects in the left hemisphere (average Talairach coordinates ± SEM: -55 ± 1 , -8 ± 1 , 5 ± 1) and in 6/8 subjects in the right hemisphere (63 ± 1 , -9 ± 1 , 4 ± 1).

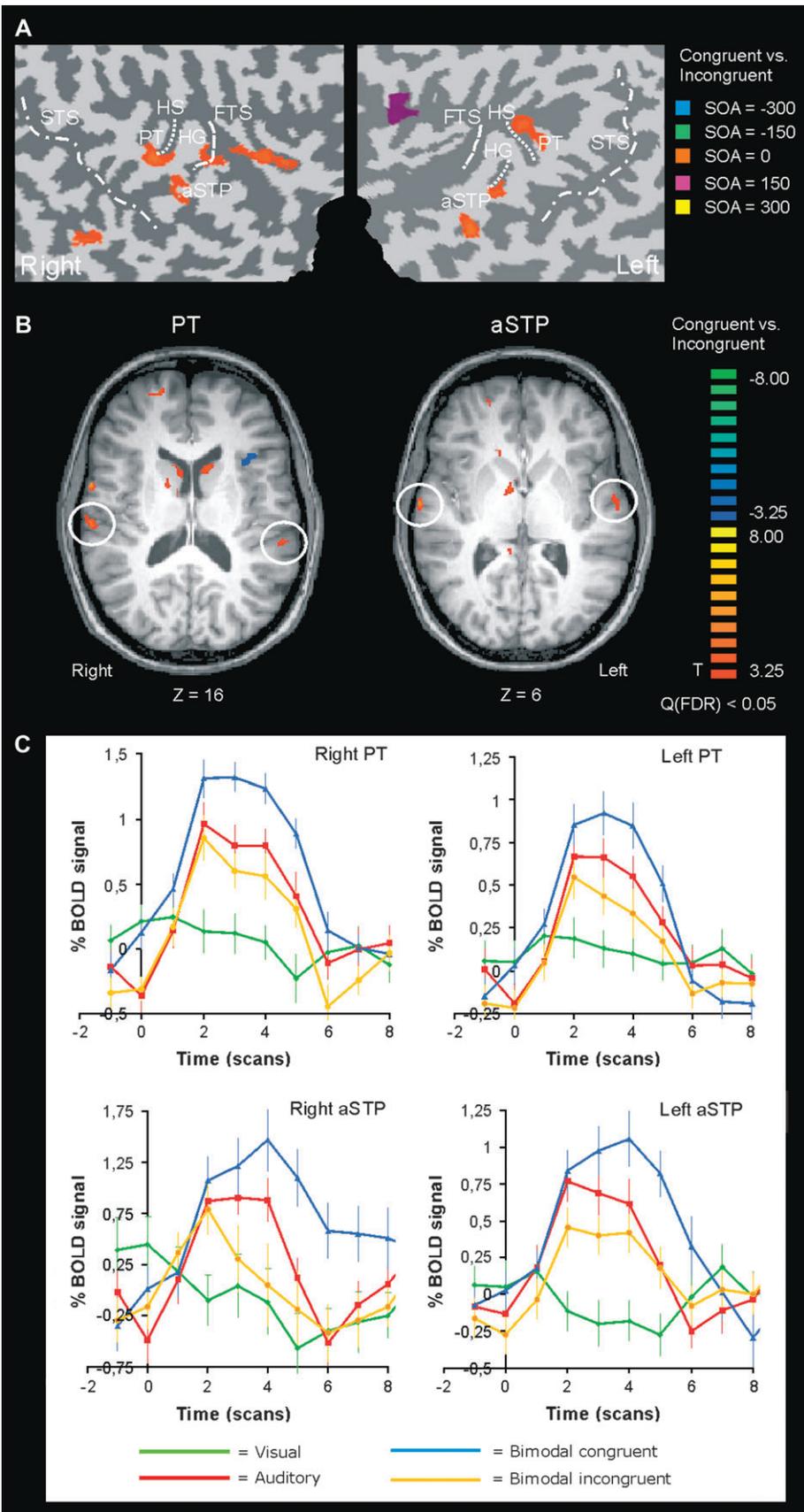
The averaged time courses of the fMRI response during bimodal stimulation at the different SOAs in PT and aSTP are shown in Figure 4B. In the PT bilaterally and left aSTP, the time courses indicate that the observed interaction was explained by a congruency effect (congruent > incongruent) that was only present at synchronous presentation (most clearly visible in the difference plots, Fig. 4B, right column). In addition to the congruency effect at SOA = 0, the congruency effect was reversed (incongruent > congruent) for SOA = -150 in the right aSTP. These observations were confirmed by ROI analyses of the congruency contrast (congruent > incongruent): left PT SOA = 0 ($P < 0.005$), all other SOAs ($P > 0.1$); right PT SOA = 0 ($P < 0.001$), all other SOAs ($P > 0.1$); left aSTP SOA = 0 ($P < 0.001$), all other SOAs ($P > 0.1$); right aSTP SOA = 0 ($P < 0.001$), SOA = -150 (incongruent > congruent, $P < 0.05$), all other SOAs ($P > 0.01$).

Figure 5 shows the response patterns in the PT and aSTP in more detail (ROIs selected by Congruency × SOA interaction at $q[\text{FDR}] < 0.05$). The bar graphs show fMRI response levels during unimodal and synchronous bimodal stimulation averaged over subjects. The PT (Fig. 5A) showed an auditory-specific

unimodal response (auditory vs. visual: $t_7 = 6.6$, $P < 0.001$ [left]; $t_7 = 5.3$, $P < 0.001$ [right]) and a strong preference for congruent as compared with incongruent letter-sound pairs (congruent vs. incongruent: $t_7 = 2.9$, $P < 0.05$ [left]; $t_7 = 2.3$, $P < 0.05$ [right]). This response pattern in the PT is a replication of the effects reported in our previous study. The aSTP (Fig. 5B) also showed an auditory-specific response pattern (auditory vs. visual: $t_7 = 4.3$, $P < 0.005$ [left]; $t_7 = 2.5$, $P < 0.05$ [right]), the congruency effect was only significant in the left hemisphere (congruent vs. incongruent: $t_7 = 3.8$, $P < 0.01$ [left]; $t_7 = 1.5$, $P > 0.1$ [right]).

To examine the effect of SOA on multisensory integration, individual MSI values for congruent and incongruent stimuli were plotted against SOA (Fig. 5, line graphs). MSI was quantified by calculating the bimodal response (AV, separately for AV congruent and AV incongruent) relative to the most effective unimodal response ($[A, V]_{\text{max}}$) in each individual subject ($((\text{AV} - [A, V]_{\text{max}}) / [A, V]_{\text{max}}) \times 100\%$, see Materials and Methods). Therefore, the terms response enhancement (positive interaction) and response depression (negative interaction) in the following refer to the bimodal response relative to the most effective unimodal response (and not relative to the baseline response). In accordance to Figure 4B, Figure 5A reveals that in the PT, the difference in MSI produced by congruent (response enhancement) and incongruent (response depression) stimulus pairs was only observed for synchronous presentation. The same effect of SOA on MSI was demonstrated for the aSTP (Fig. 5B), although in this region the congruency effect at SOA = 0 was mainly due to an enhancement for congruent stimuli, without a response depression for incongruent stimuli. As already indicated by the time courses (Fig. 4B), an interesting different effect of SOA was observed in the right aSTP (Fig. 5B). In this region, the congruency effect at SOA = 0 (congruent > incongruent) was reversed at SOA = -150 (incongruent > congruent). The response depression at this SOA was only present for congruent stimuli, indicating that the response evoked by a speech sound in this region is weaker when preceded by a visual letter of the same identity, but not when preceded by a different visual letter.

Figure 3. Results of the congruency contrast (congruent > incongruent) at all SOAs. (A) Multisubject GLM1 map of the congruency effects at different SOAs, created from cortex-based aligned functional data. Maps are shown on the flattened cortical sheet of the superior temporal lobe of the individual brain used as target for the cortex-based alignment. A congruency effect in superior temporal cortex was only found for synchronous stimuli (SOA = 0), in PT and aSTP. (B) Multisubject GLM1 map of the congruency effect at SOA = 0 shown in transversal slices. A significant congruency effect in superior temporal cortex was observed in PT and aSTP bilaterally (white circles). (C) Averaged time courses of the BOLD response (in percent signal change) in the PT and aSTP during unimodal (visual, green lines; auditory, red lines) and bimodal (congruent, blue lines; incongruent, yellow lines) synchronous stimulation periods. Error bars indicate SEM. HS, Heschl's Sulcus; HG, Heschl's gyrus; FTS, first transverse temporal sulcus.



Superior Temporal Sulcus

The interaction analysis (SOA \times Congruency) did not reveal regions in the STS. The STS has been reported to be involved in letter-speech sound integration (Raij and others 2000; Hashimoto and Sakai 2004; Van Atteveldt and others 2004) and in integration of other types of complex audiovisual information (see e.g., Beauchamp 2005a). We explored the effect of SOA in the STS using the max criterion (the conjunction of [bimodal $>$ unimodal \cap unimodal $>$ baseline], see Materials and Methods) at all SOAs. Figure 6A shows the result of the max criterion analysis at SOA = 0, which revealed a region in left STS (see also Table 1). Note that this map corresponds to the regions shown in Figure 2A, lower right (intersection auditory \cap visual), for which it is also true that the response to bimodal stimulation is stronger than the response to unimodal stimulation. From the regions shown in this intersection map, only the left STS region passed this additional criterion. The response pattern in the left STS, shown by the BOLD response time courses in Figure 6A, is a replication of the pattern found in our previous study (Van Atteveldt and others 2004).

Figure 6C shows fMRI response levels for the unimodal and bimodal synchronous conditions in the left STS averaged over subjects (bar graphs) and the corresponding MSI values (line graph). The response pattern shown in the bar graph indicates that the enhanced response for bimodal stimulation was significant across subjects (congruent vs. auditory: $t_7 = 3.2$, $P < 0.05$; congruent vs. visual: $t_7 = 3.4$, $P < 0.01$; incongruent vs. auditory: $t_7 = 3.3$, $P < 0.05$; incongruent vs. visual: $t_7 = 3.7$, $P < 0.01$). In contrast to the auditory-specific response pattern in the PT and aSTP, the STS showed a heteromodal response pattern (auditory vs. visual, $t_7 = 0.9$, $P = 0.4$), indicating multisensory convergence. In addition, no congruency effect was observed in the STS (congruent vs. incongruent, $t_7 = 0.5$, $P = 0.6$).

The max criterion analysis revealed a similar region in left STS for all SOAs (Fig. 6B), which indicates that a temporal offset between letters and speech sounds did not have an effect in the STS similar to that demonstrated for the auditory association cortex. This observation was confirmed by the MSIs (Fig. 6C, line graphs): significant positive MSIs were observed for both bimodal conditions at all SOAs (except for at SOA = -150 [congruent] and at SOA = 150 [incongruent]).

Discussion

The principle aim of the present study was to elucidate the effect of temporal asynchrony on the neural integration of letters and speech sounds. We manipulated both the temporal relation (SOA) and content congruency (same/different identity) between letters and speech sounds within the same experimental design. Of particular interest for the present study are regions showing an interaction between SOA and content congruency when causing fMRI responses to letter-sound pairs because such regions provide direct evidence for an influence of temporal relation on the neural binding of letters and speech sounds. The results clearly demonstrate that temporal relation and information content interact when causing fMRI responses to letter-speech sound pairs in anterior and posterior auditory association cortex (aSTP and PT), but not in the STS.

Auditory Association Cortex

One highly interesting observation is that temporal synchrony is a prerequisite for the occurrence of multisensory integration of

letters and speech sounds in the posterior part of the auditory association cortex, the PT. The posterior part of the auditory cortex has been shown to play an important role in speech perception (e.g., Zatorre and others 1992; Jäncke and others 2002; Buchsbaum and others 2005), and more specifically in the integration of written and spoken language (Nakada and others 2001; Van Atteveldt and others 2004). As is shown in Figure 5A (line graphs), both the magnitude of response enhancement during congruent stimulation as well as the magnitude of response depression during incongruent stimulation rapidly declined with temporal asynchrony. This observation implies that temporal correspondence overrules information content as binding factor, which is in accordance with predictions made by the time-window-of-integration model for multisensory integration (Colonius and Diederich 2004; Diederich and Colonius 2004). This model assumes that the time interval between the unisensory inputs acts like a filter by determining the probability of interaction. Other factors such as spatial configuration of the stimuli, and possibly also information content as suggested by the present results, have a subsequent role in determining the amount and direction (enhancement or depression) of interaction, once the temporal filter has been passed successfully. In the context of the present study, the dominance of temporal synchrony as determining factor for integration is a highly interesting finding since we studied multisensory associations that were initially only related by information content. This finding therefore supports the idea that basic neural integration rules apply to the binding of overlearned multisensory associations that are not naturally related.

Temporal relation and content congruency also interacted in the auditory association cortex anterior to the primary auditory cortex (aSTP). However, the effect of SOA in aSTP shows subtle differences from the effects observed in PT (line graphs in Fig. 5). In the left aSTP, the congruency effect for synchronous stimuli is mainly due to an enhancement for congruent stimuli, without a depression for incongruent stimuli. Interestingly, in the right aSTP, the congruency effect was reversed when the visual stimulus preceded the auditory stimulus by 150 ms (SOA = -150, incongruent $>$ congruent). At this SOA, the response to congruent bimodal stimuli is weaker than the response to speech sounds presented alone (response depression), whereas the response to incongruent stimuli is not different from the unimodal response. The reduced fMRI response to speech sounds preceded by visually presented letters of the same identity might be explained by a cross-modal repetition suppression (Henson 2003) or functional magnetic resonance (fMR) - adaptation (Grill-Spector and Malach 2001) effect. Reduction of the fMR signal by repeated presentation of a single stimulus has been demonstrated within modalities and is thought to reflect neuronal adaptation. Although this interpretation is speculative at this point, fMR-adaptation designs may provide a way to gain insight in the functional characteristics of connections between different sensory systems in future research. By specifically tagging neuronal populations that are cross-modally activated, detailed investigation of the functional properties of these intersensory connections will be possible.

The demonstrated effects of congruency in the auditory association cortex might alternatively be explained in terms of attention. Because we used a block design, subjects know from the first stimulus of a block whether all subsequent stimuli will be congruent or incongruent. This might lead to increased attention to the stimuli in the congruent blocks and decreased

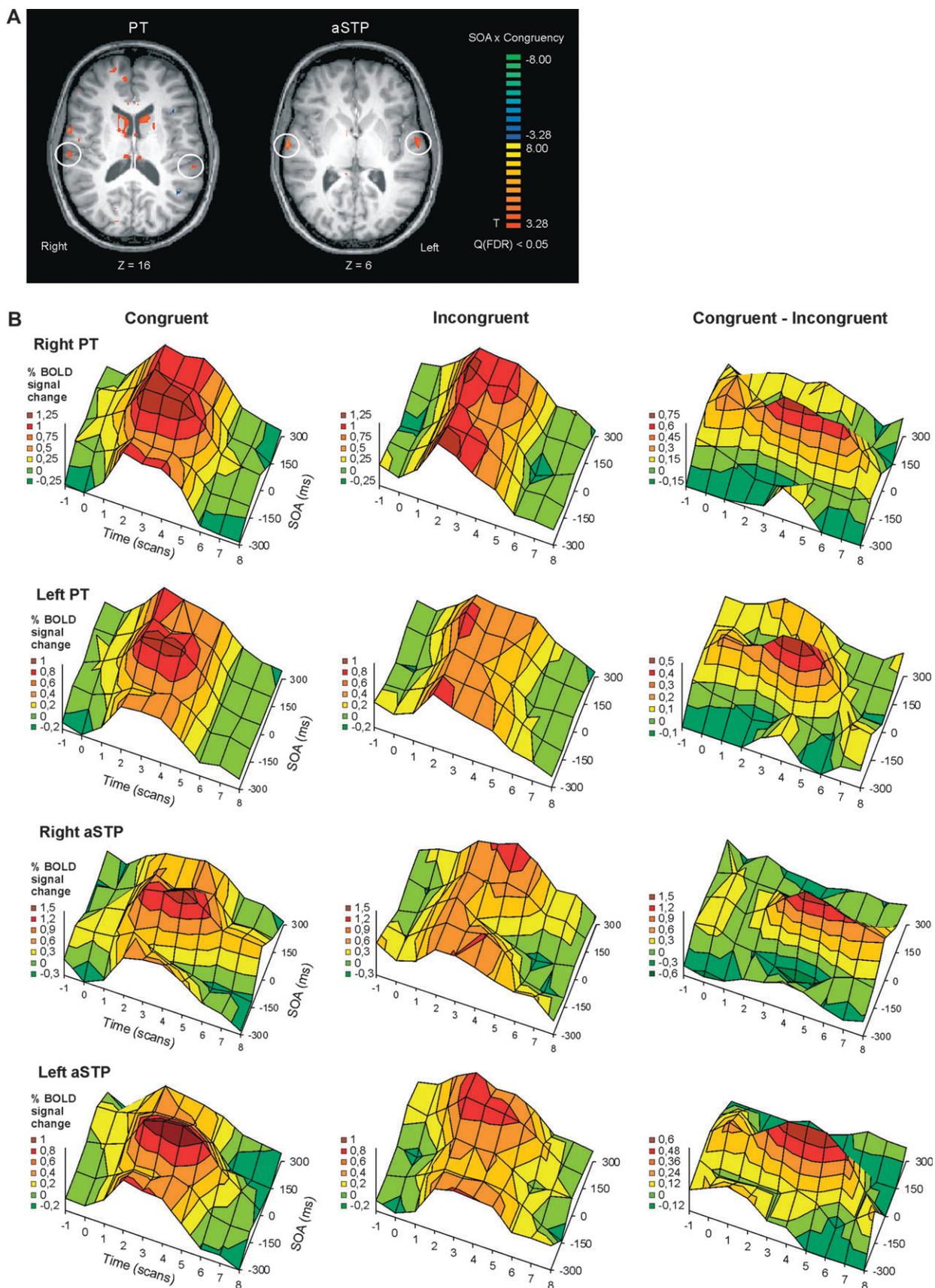


Figure 4. Results of the interaction analysis (Congruency \times SOA). (A) Multisubject factorial GLM2 showing the interaction of SOA and Congruency in transversal slices. (B) Averaged time courses of the BOLD response (in percent signal change, indicated by the color coding on the y axis) in the PT and aSTP bilaterally for all SOAs (plotted on the z axis), plotted separately for congruent (left) and incongruent (middle) bimodal stimulation, and the difference between congruent and incongruent (right). The stimulation starts at time = 0.

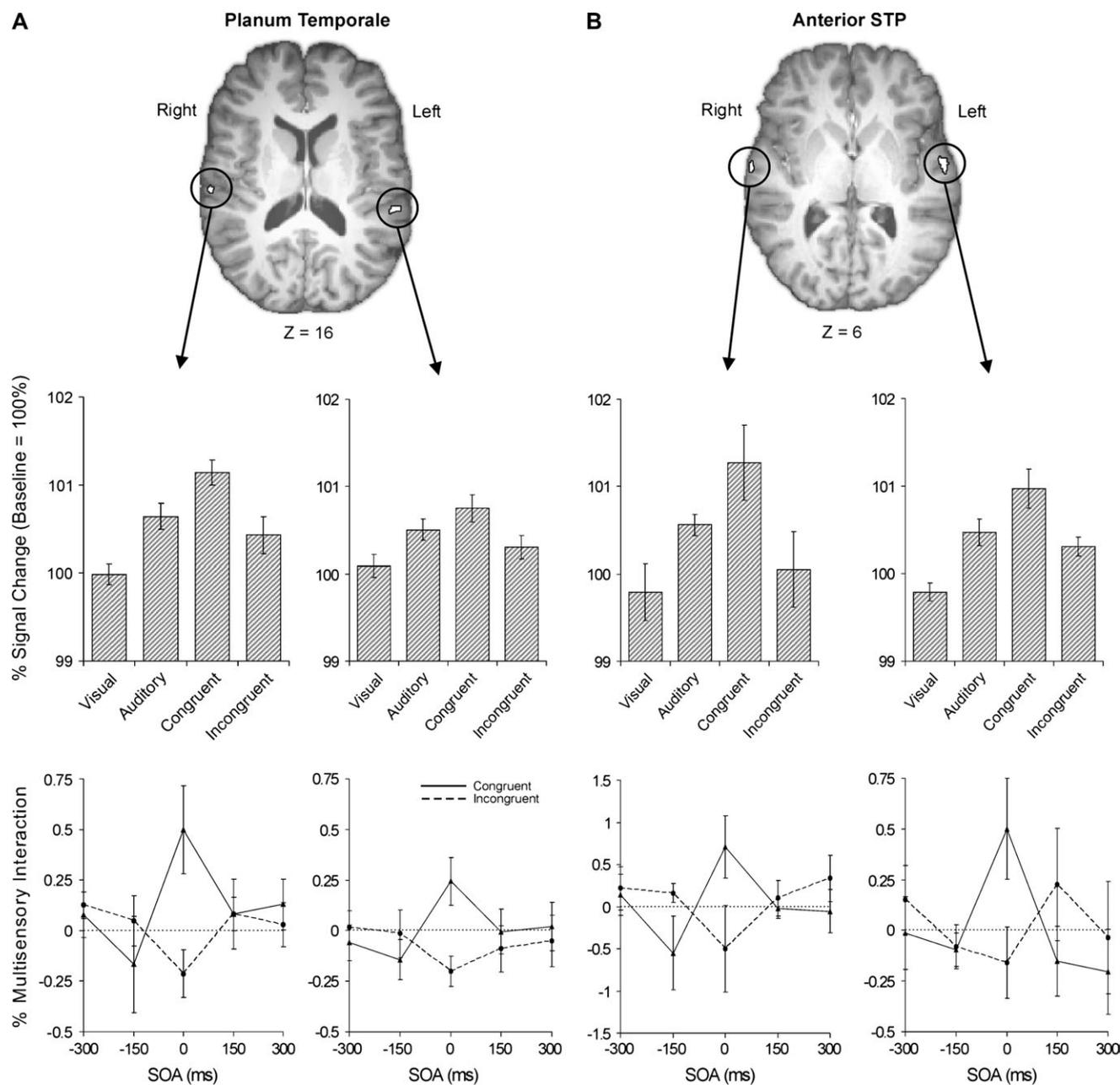
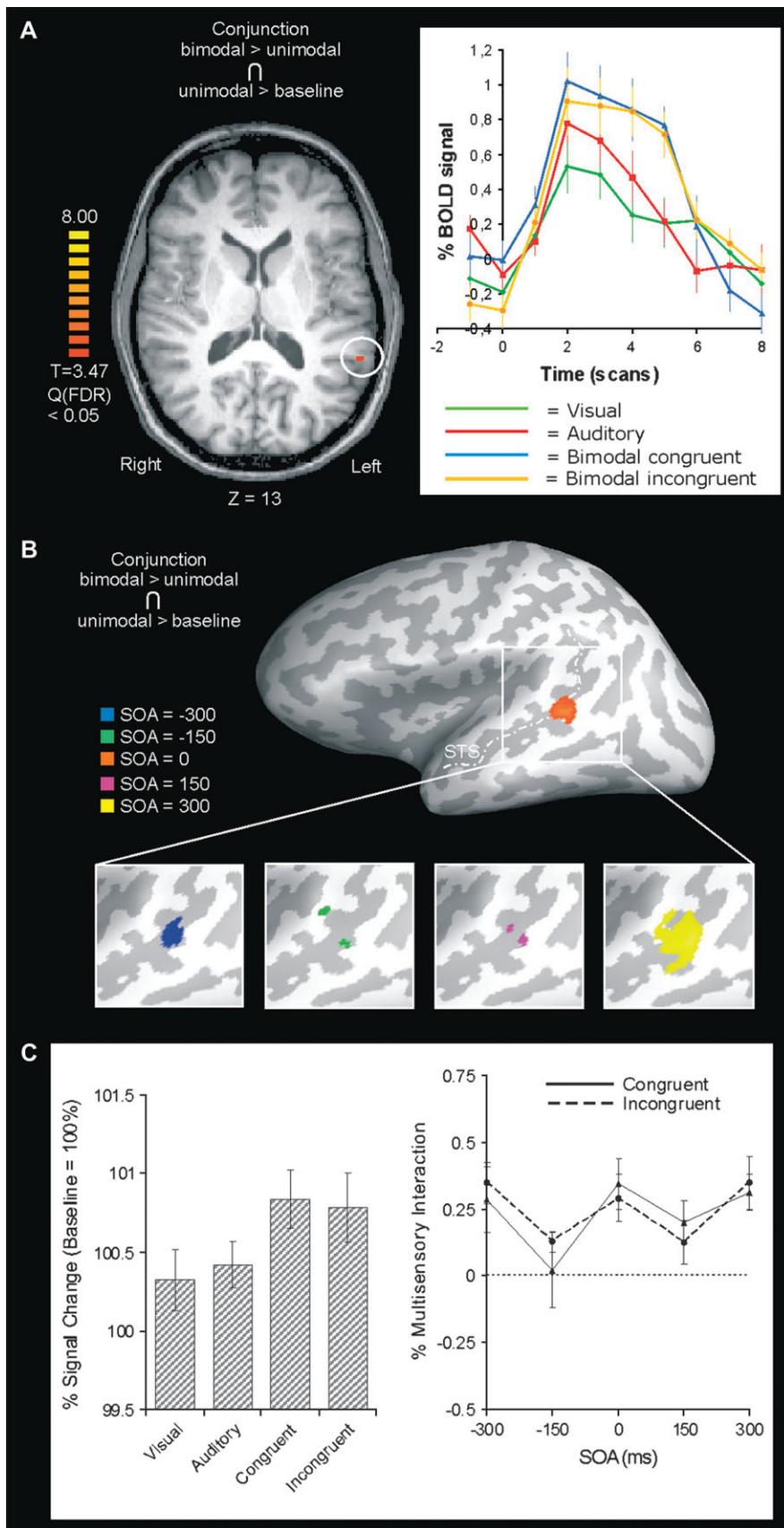


Figure 5. Response patterns and effect of SOA on MSIs in PT (A) and aSTP (B). Bar graphs: MR signal levels for the unimodal and bimodal synchronous conditions, averaged over subjects (error bars indicate SEM). For each subject individually, the average MR signal during fixation periods (baseline) was set at 100%. Line graphs: MSI for congruent (solid lines) and incongruent (dashed lines) bimodal stimulation plotted as a function of SOA. MSI is defined as the bimodal response as percentage of the most effective unimodal response and was calculated for each subject at each condition using the MR signal values plotted in the corresponding bar graphs. Error bars indicate variability across subjects (SEM).

attention in the incongruent blocks, resulting in the observed response enhancement and depression. However, considering the high specificity of the congruency effect to focal regions in auditory association cortex, we think, an explanation in terms of

a general attention mechanism is unlikely because this would predict an effect of congruency to be more widespread in the auditory cortex and to also include attention areas. Furthermore, attention alone cannot explain why the congruency

Figure 6. Results of max criterion analysis (bimodal > unimodal) \cap (unimodal > baseline). (A) Results of max criterion analysis at SOA = 0 shown in a transversal slice and the corresponding averaged time course of the BOLD response (visual, green lines; auditory, red lines; congruent, blue lines; incongruent, yellow lines). (B) Results of the max criterion analysis for all SOAs performed on cortex-based aligned data and shown on the cortical surface of the individual brain used as target for the alignment. (C) Bar graph: MR signal levels for the unimodal and bimodal synchronous conditions in the left STS, averaged over subjects (error bars indicate SEM). Line graph: MSI for congruent (solid lines) and incongruent (dashed lines) bimodal stimulation plotted as a function of SOA. MSI is defined as the bimodal response as percentage of the most effective unimodal response and was calculated for each subject at each condition using the MR signal values plotted in the corresponding bar graphs. Error bars indicate variability across subjects (SEM).



effect disappears, or even inverts (as observed in the right aSTP), when letters and sounds are asynchronously presented. Therefore, it seems plausible that the congruency effects in the auditory association cortex reflect (the result of) cross-modal integration. This is strongly supported by the characterization of multisensory integration by response enhancement and suppression in nonhuman electrophysiological studies (for a review, see Stein and others 2004) and other human fMRI studies (Calvert and others 2000; Saito and others 2005).

The observed MSI effects in the auditory association cortex suggest that speech processing is influenced by visual orthographic information in focal regions anteriorly as well as posteriorly from the primary auditory cortex. Although the functional role of the anterior and posterior auditory processing streams is still under debate (Scott 2005), (nonspatial) speech processing is reported in anterior as well as posterior superior temporal cortices (Arnott and others 2004). The different temporal profile of MSIs for both regions in the present study may suggest involvement in different aspects of letter-speech sound integration. The presumed cross-modal repetition suppression observed in the right aSTP may suggest a role in associating the exact identity of letters and speech sounds (the “what” pathway), whereas the PT may be involved in the “how” pathway, which is thought to be involved in sensory motor integration of speech information (Buchsbaum and others 2005; Scott 2005). Consistent with the view of the PT as “computational hub” (Griffiths and Warren 2002) or sensory motor interface (Buchsbaum and others 2005; Scott 2005), the PT might link sensory representations of letters and speech sounds with motor representations involved in speaking (Wilson and others 2004) and writing (Longcamp and others 2003). This view is supported by the activation of premotor cortex by the unimodally presented letters and speech sounds (Fig. 2).

Superior Temporal Sulcus

We found a heteromodal region in the left STS in which the bimodal response exceeded both unimodal responses, consistent with our previous study and with the assumed role of the STS in integration of letters and speech sounds (Raij and others 2000; Hashimoto and Sakai 2004; Van Atteveldt and others 2004) and other types of audiovisual identity information (Calvert 2001; Beauchamp and others 2004; Amedi and others 2005; Beauchamp 2005a). Congruent and incongruent bimodal stimuli both evoked enhanced responses in the STS, which may seem unexpected considering the assumed integrative function. A possible explanation is that if congruency is determined in the STS, both congruent and incongruent combinations need computation and might therefore both lead to increased neural activity. This is in accordance to the fMRI study on complex audiovisual objects by Beauchamp and others (2004) who also did not find a significant effect of congruency in the STS. In contrast to the present findings, Calvert and others (2000) report an enhanced fMRI response for congruent and a depressed fMRI response for incongruent audiovisual speech. Other than design differences, this discrepancy might be related to the different nature and learning of audiovisual speech and letter-sound combinations (see also Van Atteveldt and others 2004). Whereas audiovisual speech occurs naturally and is learned early and implicitly (Kuhl and Meltzoff 1982), letters are artificial and have to be associated with speech sounds by explicit instruction during literacy acquisition (Lieberman

1992). These differences might cause different computational demands during audiovisual integration in the STS. Using magnetoencephalography (MEG), Raij and others (2000) found differential interactions (although both negative) for congruent and incongruent audiovisual letters in the STS, which may seem contradictory to this interpretation. However, regarding the limited spatial resolution of MEG, the congruency effect in the study of Raij and others may also have originated from slightly more superior temporal cortex, corresponding to the regions showing congruency effects in the present study (PT and aSTP).

Compared to the auditory association cortex, integration in the STS is less dependent on temporal synchrony (Fig. 6), which is consistent with previous neuroimaging findings (Olson and others 2002). Furthermore, the integration of audiovisual speech, which is thought to depend on integration in the STS (e.g., Calvert and others 2000), has shown to be relatively unaffected by temporal disparity (Massaro and Cohen 1993; Massaro and others 1996; Munhall and others 1996). Although integration in the left STS occurs within a wide temporal window in the present study, it appears to be least effective when the temporal offset between the visual and auditory stimuli is small (see Fig. 6C).

Implications for the Neural Mechanism of Letter-Speech Sound Integration

Based on our findings, we propose the following neural mechanism of letter-speech sound integration (see also Van Atteveldt and others 2004). Speech sounds are likely to be primarily represented and processed in the PT (Hickok and Poeppel 2000; Griffiths and Warren 2002). The next processing level, the STS, also receives visual information and integrates both inputs within a broad range of SOAs. Depending on the temporal relationship between the inputs from both modalities, feedback regarding identity congruency is sent to the auditory association cortex, resulting in the observed temporal profiles of MSI there. A wider temporal window of integration in the STS enables a more flexible use of learned associations. It seems therefore plausible that the observed temporal windows for integration will be influenced by top-down strategic control when a task is introduced (Dijkstra and others 1989). However, in the passive viewing and listening situation of the present study, basic rules of temporal proximity seem to apply to the automatic binding of letters and speech sounds, and feedback to the PT and left aSTP seems only to be provided when the stimuli are presented in temporal synchrony. Feedback to the right aSTP is also sent at short negative SOAs and has the reversed effect on speech sound processing (depression for congruent subsequent stimuli), which may reflect cross-modal repetition suppression or adaptation. Furthermore, our data suggest that the STS sends feedback to aSTP and PT with different purposes: aSTP for identification processes and PT for processes requiring sensory motor integration. The PT may subsequently project to frontal and parietal regions involved in speech production and writing.

The response patterns and effects of temporal asynchrony observed in the auditory association cortex bears resemblance to those demonstrated for single multisensory neurons across brain areas and animal species (Meredith and others 1987; Stein and Wallace 1996; Wallace and others 1996). This similarity suggests that multisensory neurons with similar properties exist in the human auditory association cortex and thus that

integration may take place directly there. Support for this suggestion is provided by the recent demonstration of integration of multisensory inputs in the auditory association cortex in macaques (Schroeder and others 2001; Schroeder and Foxe 2002), which has recently been demonstrated to be strongest for temporally coincident stimuli (Kayser and others 2005). However, laminar input profiles indicated that visual input in the auditory cortex probably reflects feedback rather than direct input, possibly originating from the superior temporal polysensory area (Schroeder and Foxe 2005), an area in the macaque that may correspond to the human multisensory STS (Beauchamp 2005a). Furthermore, the PT and aSTP do not respond to visual unimodal stimulation (Figs 3C and 5), whereas the STS shows multisensory convergence (Fig. 6). Therefore, we think it is more plausible that the STS serves as an extra processing level where associations between letters and speech sounds are established, as was also indicated by our previous fMRI study (Van Atteveldt and others 2004).

Whereas audiovisual speech integration is known to be relatively unaffected by temporal asynchrony (Massaro and Cohen 1993; Massaro and others 1996; Munhall and others 1996; Munhall and Vatikiotis-Bateson 2004), the present study shows more stringent temporal constraints for the integration of letters and speech sounds. This apparent discrepancy may be explained by the fact that in audiovisual speech, the visual and auditory inputs share more features, for example, time-varying aspects such as frequency amplitude information (Munhall and others 1996; Calvert and others 1998; Munhall and Vatikiotis-Bateson 1998; Amedi and others 2005). Because letters and speech sounds lack these naturally corresponding features, it is tentative to assume that simultaneous onset is more critical for their integration. This idea bears resemblance to the finding of Dixon and Spitz (1980) that asynchrony of audiovisual information with less concordant time-varying information (a hammer hitting a peg) is more easily detected than that of audiovisual speech.

Conclusions

In summary, multisensory integration of letters and speech sounds in the human auditory association cortex showed a strong dependency on the relative timing of the inputs. The critical role of input timing on multisensory integration has been demonstrated before at the neuronal level for naturally related visual and auditory signals. This similarity suggests that basic neural integration rules apply to the binding of multisensory information that is not naturally related but overlearned during literacy acquisition. However, the mechanism by which the temporal constraints are effected may differ, that is, the temporal windows in the auditory association cortex observed in the present study may be the result of feedback from the STS.

Notes

This work was supported by grant 608/002/2005 of the Dutch Board of Health Care Insurance (College voor Zorgverzekeringen) awarded to LB. We thank Peter Hagoort for providing access to the facilities of the F.C. Donders Centre and Paul Gaalman for his technical assistance. *Conflict of Interest:* None declared.

Address correspondence to Nienke M. van Atteveldt, Department of Cognitive Neuroscience, University of Maastricht, P.O. Box 616, 6200

MD Maastricht, The Netherlands. Email: N.vanAtteveldt@psychology.unimaas.nl.

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.
- Arnott SR, Binns MA, Grady CL, Alain C. 2004. Assessing the auditory dual-pathway model in humans. *Neuroimage* 22:401–408.
- Beauchamp M. 2005a. See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:1–9.
- Beauchamp M. 2005b. Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113.
- 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.
- Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4241.
- Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Shane Kippenhan J, Faith Berman K. 2005. Reading, hearing, and the planum temporale. *Neuroimage* 24:444–454.
- Calvert GA. 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Brammer MJ, Iversen SD. 1998. Crossmodal identification. *Trends Cogn Sci* 2:247–253.
- 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.
- Colonus H, Diederich A. 2004. Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci* 16:1000–1009.
- Diederich A, Colonus H. 2004. Modeling the time-course of multisensory interaction in manual and saccadic responses. In: Calvert GA, Spence C, Stein BE, editors. *The handbook of multisensory processes*. Cambridge, MA: The MIT Press. p 395–408.
- Dijkstra A, Schreuder R, Frauenfelder UH. 1989. Grapheme context effects on phonemic processing. *Lang Speech* 32:89–108.
- Dixon NF, Spitz L. 1980. The detection of auditory visual desynchrony. *Perception* 9:719–721.
- Ehri LC. 2005. Development of sight word reading: phases and findings. In: Snowling MJ, Hulme C, editors. *The science of reading: a handbook*. Oxford: Blackwell Publishing. p 135–154.
- Flowers DL, Jones K, Noble K, VanMeter J, Zeffiro TA, Wood FB, Eden GF. 2004. Attention to single letters activates left extrastriate cortex. *Neuroimage* 21:829–839.
- Genovese C, Lazar N, Nichols T. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol* 107:293–321.
- Hashimoto R, Sakai KL. 2004. Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron* 42:311–322.
- Henson R. 2003. Neuroimaging studies of priming. *Prog Neurobiol* 70:53–81.
- Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131–138.
- Jäncke L, Wüstenberg T, Scheich H, Heinze HJ. 2002. Phonetic perception and the temporal cortex. *Neuroimage* 15:733–746.
- Kayser C, Petkov C, Augath M, Logothetis NK. 2005. Integration of touch and sound in auditory cortex. *Neuron* 48:373–384.
- Kuhl PK, Meltzoff AN. 1982. The bimodal perception of speech in infancy. *Science* 218:1138–1141.

- Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT. 2004. Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res* 158:405–414.
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE. 2005. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289–297.
- Liberman AM. 1992. The relation of speech to reading and writing. In: Frost R, Katz L, editors. *Orthography, phonology, morphology and meaning*. Amsterdam, The Netherlands: Elsevier Science Publishers BV. p 167–178.
- Longcamp M, Anton JL, Roth M, Velay JL. 2003. Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage* 19:1492–1500.
- Massaro DW, Cohen MM. 1993. Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Commun* 13:127–134.
- Massaro DW, Cohen MM, Smeele PM. 1996. Perception of asynchronous and conflicting visual and auditory speech. *J Acoust Soc Am* 100:1777–1786.
- Matsuo K, Kato C, Sumiyoshi C, Toma K, Thuy DHD, Moriya T, Fukuyama H, Nakai T. 2003. Discrimination of Exner's area and the frontal eye field in humans—functional magnetic resonance imaging during language and saccade tasks. *Neurosci Lett* 340:13–16.
- Meredith MA, Nemitz JW, Stein BE. 1987. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215–3229.
- Meredith MA, Stein BE. 1983. Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389–391.
- Munhall K, Gribble P, Sacco L, Ward M. 1996. Temporal constraints on the McGurk effect. *Percept Psychophys* 58:351–362.
- Munhall K, Vatikiotis-Bateson E. 1998. The moving face during speech communication. In: Campbell R, Dodd B, Burnham D, editors. *Hearing by eye II: The psychology of speechreading and audio visual speech*. London, UK: Psychology Press. p 123–139.
- Munhall K, Vatikiotis-Bateson E. 2004. Spatial and temporal constraints on audiovisual speech perception. In: Calvert GA, Spence C, Stein BE, editors. *The handbook of multisensory processes*. Cambridge, MA: The MIT Press. p 177–188.
- Nakada T, Fujii Y, Yoneoka Y, Kwee IL. 2001. Planum temporale: where spoken and written language meet. *Eur Neurol* 46:121–125.
- Olson IR, Christopher Gatenby J, Gore JC. 2002. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cogn Brain Res* 14:129–138.
- Pourtois G, de Gelder B. 2002. Semantic factors influence multisensory pairing: a transcranial magnetic stimulation study. *Neuroreport* 13:1567–1573.
- Raij T, Uutela K, Hari R. 2000. Audiovisual integration of letters in the human brain. *Neuron* 28:617–625.
- 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.
- Schroeder CE, Foxe JJ. 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn Brain Res* 14:187–198.
- Schroeder CE, Foxe JJ. 2005. Multisensory contributions to low-level, 'unisensory' processing. *Curr Opin Neurobiol* 15:1–5.
- Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC. 2001. Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol* 85:1322–1327.
- Scott SK. 2005. Auditory processing—speech, space and auditory objects. *Curr Opin Neurobiol* 15:197–201.
- Stein BE, Jiang H, Stanford TR. 2004. Multisensory integration in single neurons of the midbrain. In: Calvert GA, Spence C, Stein BE, editors. *The handbook of multisensory processes*. Cambridge, MA: The MIT Press. p 243–264.
- Stein BE, Meredith MA. 1993. *The merging of the senses*. Cambridge, MA: MIT Press.
- Stein BE, Wallace MT. 1996. Comparisons of cross-modality integration in midbrain and cortex. *Prog Brain Res* 112:289–299.
- Van Atteveldt N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM. 2004. Specific reading disability (dyslexia): what have we learned in the past four decades? *J Child Psychol Psychiatry* 45:2–40.
- Wallace MT, Wilkinson LK, Stein BE. 1996. Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246–1266.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M. 2004. Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G. 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034–1043.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.