Neuron, Vol. 43, 271–282, July 22, 2004, Copyright ©2004 by Cell Press

Integration of Letters and Speech Sounds in the Human Brain

Nienke van Atteveldt,[1,]* Elia Formisano,[1]
Rainer Goebel,[1,2] and Leo Blomert[1]
[1]Faculty of Psychology
Department of Cognitive Neuroscience
University of Maastricht
P.O. Box 616
6200 MD Maastricht
The Netherlands
[2]F.C. Donders Centre for Cognitive Neuroimaging
P.O. Box 9101
6500 HB Nijmegen
The Netherlands

Summary

Most people acquire literacy skills with remarkable ease, even though the human brain is not evolutionarily adapted to this relatively new cultural phenomenon. Associations between letters and speech sounds form the basis of reading in alphabetic scripts. We investigated the functional neuroanatomy of the integration of letters and speech sounds using functional magnetic resonance imaging (fMRI). Letters and speech sounds were presented unimodally and bimodally in congruent or incongruent combinations. Analysis of single-subject data and group data aligned on the basis of individual cortical anatomy revealed that letters and speech sounds are integrated in heteromodal superior temporal cortex. Interestingly, responses to speech sounds in a modality-specific region of the early auditory cortex were modified by simultaneously presented letters. These results suggest that efficient processing of culturally defined associations between letters and speech sounds relies on neural mechanisms similar to those naturally evolved for integrating audiovisual speech.

Introduction

Reading is essential to social and economic success in the present technological society (National Reading Council, 1998). In contrast to spoken language, which is a product of biological evolution, reading and writing are cultural inventions from the last few thousand years and are only relevant for most people since a few hundred years (Liberman, 1992). An intriguing question is, therefore, how it is possible that most people acquire literacy skills with such remarkable ease even though a naturally evolved brain mechanism for reading is unlikely to exist. An interesting hypothesis is that evolutionarily adapted brain systems for spoken language provide a neural foundation for reading ability, which is illustrated by the low literacy levels in deaf people (Perfetti and Sandak, 2000).

Nowadays most written languages are speech-based alphabetic scripts, in which speech sound units (phonemes) are represented by visual symbols (letters, or graphemes). Learning the correspondences between letters and speech sounds of a language is therefore a crucial step in reading acquisition, failure of which is thought to account for reading problems in developmental dyslexia (Frith, 1985). However, in the normal situation, letter-speech sound associations are learned and used with high efficiency. At least 90% of school children learn the letter-sound correspondences without exceptional effort within a few months (Blomert, 2002), which is a remarkable achievement, since our brains are not phylogenetically adapted to the requirements for acquiring written language.

Associations between sensory events in different modalities can either be defined by natural relations (e.g., the shape and sound of a natural object) or by more artificial relations. In contrast to the culturally defined associations between letters and speech sounds (Raij et al., 2000), lip reading is based on naturally developed associations of speech with visual information (Paulesu et al., 2003). Therefore, it seems a plausible assumption that the perception of speech and the inherently linked lip movements (hereafter referred to as *audiovisual speech*) emerged simultaneously during evolution, shaping the brain for integrating this audiovisual information.

At the behavioral level, it has been reported that speech perception can be influenced both by lip movements and by letters. Sumby and Pollack (1954) showed that lip reading can improve speech perception, especially in situations when the auditory input is degraded. More extremely, lip reading can also *change* the auditory speech percept, as is shown in the McGurk effect (McGurk and MacDonald, 1976). Improvement of speech perception by simultaneous presentation of print has been demonstrated at the level of words (Frost et al., 1988) and syllables (Massaro et al., 1988). Dijkstra and colleagues reported facilitation and inhibition effects on auditorily presented phoneme identity decisions by congruent and incongruent letter primes, respectively, suggesting activation of phoneme representations by letters (Dijkstra et al., 1989).

A neural mechanism for the integration of audiovisual speech has been suggested by Calvert and colleagues (Calvert et al., 1999, 2000) and supported by other neuroimaging findings on audiovisual speech perception (Sams et al., 1991; Sekiyama et al., 2003; Wright et al., 2003) and lip reading (Calvert et al., 1997; Calvert and Campbell, 2003; Paulesu et al., 2003). Results of these studies suggest that the perceptual gain experienced when perceiving multimodal speech is accomplished by enhancement of the neural activity in the relevant sensory cortices. The left posterior superior temporal sulcus (STS) has been advanced as the heteromodal site that integrates visual and auditory speech information and modulates the modality-specific cortices by back projections (Calvert et al., 1999, 2000). Modality-specific regions involved in this mechanism are the visual motion processing area V5 and auditory association areas in superior temporal cortex. In addition to this interplay between STS and sensory cortices, frontal and

*Correspondence: n.vanatteveldt@psychology.unimaas.nl parietal regions seem to be involved, although activation of these regions is less consistent between the different studies. Interestingly, the involvement of the left posterior STS in the integration of auditory and visual nonlinguistic information has also been reported recently (Beauchamp et al., 2004; Calvert et al., 2001). These results suggest that the STS has a more general role in the integration of cross-modal identity information.

As opposed to the integration of lip movements and speech sounds, the neural mechanism that mediates the integration of letters and speech sounds is less clear. As yet, the only neuroimaging study that directly investigated this issue using multimodal stimulation was conducted by Raij and colleagues, who measured magnetic brain responses to unimodally and bimodally presented letters and speech sounds using magnetoencephalography (MEG) (Raij et al., 2000). They report a sequence of processes by which letters and speech sounds may be integrated, starting with modality-specific activation (60–120 ms after stimulus onset) in corresponding sensory cortices, followed by convergence of auditory and visual activations (around 225 ms) in lateral midtemporal cortex. Interaction of auditory and visual responses in the right temporo-occipital-parietal junction started at 280 ms; differential interaction effects for matching and nonmatching letters and speech sounds were observed in the STS at 380–540 ms. Although the time course of letter-speech sound integration provided by this study is highly informative, the spatial resolution of MEG limits the accuracy by which the exact locations of brain areas responsible for the different processes can be determined.

In the present study, we use functional magnetic resonance imaging (fMRI) at 3 Tesla to investigate the functional neuroanatomy of the integration of letters and speech sounds with high spatial resolution. Subjects passively listened to and/or viewed unimodally or bimodally presented speech sounds and letters; bimodal stimuli were either congruent or incongruent. The use of a passive perception task has been shown to be efficient by other cross-modal fMRI studies (Calvert et al., 2000; Wright et al., 2003).

The unimodal conditions were used to find brain areas responding to either letters or speech sounds (modality-specific regions) and areas responding to both (convergence regions). The bimodal conditions were used to identify regions responding more to bimodal than to unimodal stimuli (integration) and regions that differently respond to congruent and incongruent combinations of letters and speech sounds (congruency effects). We analyzed our data with different approaches. First, we searched for regions that responded stronger to bimodal (AV, audiovisual) stimulation than to each unimodal condition (V, visual; A, auditory) [$(AV > A) \cap (AV > V)$] (Beauchamp et al., 2004; Calvert et al., 1999; Hadjikhani and Roland, 1998). Second, we followed the assumption that integration is performed on converging inputs from each modality by multisensory neurons (Meredith, 2002; Raij et al., 2000), by including the constraint that integration sites should also respond significantly to both unimodal conditions in isolation [$(AV > A) \cap (AV > V) \cap A \cap V$]. Third, we determined interaction effects [$AV \neq (A + V)$] in individually defined regions of interest. Finally, we explored the effects of the relatedness of letters and speech sounds by contrasting the congruent and incongruent bimodal conditions.

Functional MRI results are typically presented as group activation maps whereby the individual data are transformed into a standard space and averaged. The high intersubject variability in brain anatomy limits the spatial accuracy of group maps produced in this way. We avoided this limitation in two ways. First, our experimental design and visualization methods allowed us to analyze and present data from individual subjects. Second, we analyzed the group data aligned on the basis of individual cortical anatomy (see Experimental Procedures). A major advantage of cortex-based intersubject alignment above more traditional methods of spatial normalization is that it increases statistical power while preserving high anatomical accuracy in the group results.

Another important aspect of the present study is that the stimuli were presented in silent intervals between subsequent volume scans (see Experimental Procedures). This method minimizes the effects of the acoustic noise produced by fast gradient switching during functional imaging on experimentally evoked auditory activation (Jäncke et al., 2002; Shah et al., 2000). This allows a highly accurate investigation of the contribution of auditory brain regions to cross-modal integration processes and more particularly the processing of letter-speech sound associations.

Results

Group Results—Unimodal Conditions

Figure 1A shows the result of the multisubject multiple linear regression analysis after cortex-based alignment of anatomical and functional data (see Experimental Procedures). The statistical map shows the activation for the unimodal predictors versus baseline [$t = 8.2$, $q(FDR) < 0.000$], color-coded for relative contribution of unimodal visual (green), unimodal auditory (red), and similar contribution of the visual and auditory predictors (yellow) to explaining the signal variance. Unimodal presentation of letters activated the lateral and inferior occipital-temporal cortex; unimodal presentation of speech sounds activated Heschl's gyrus (primary auditory cortex), and regions of the superior temporal gyrus (STG) and the STS. More posteriorly located regions of the STS and STG were activated by both unimodal conditions.

Figure 1B shows averaged time courses of the BOLD response corresponding to the regions shown in the relative contribution map (Figure 1A). Time courses are shown for three regions with different response patterns: a region in the auditory cortex (Heschl's gyrus) showing modality-specific auditory activation [unimodal auditory > unimodal visual, $q(FDR) < 0.001$], a region in posterior STS/STG showing significant response to both unimodal conditions, and a region in the visual cortex (fusiform gyrus) showing modality-specific visual activation [unimodal visual > unimodal auditory, $q(FDR) < 0.001$]. The time courses confirm that the responses in auditory and visual cortices were highly modality specific, while the posterior STS/STG showed a heteromodal response pattern. These heteromodal regions in STS and STG are candidate regions for multisensory convergence.

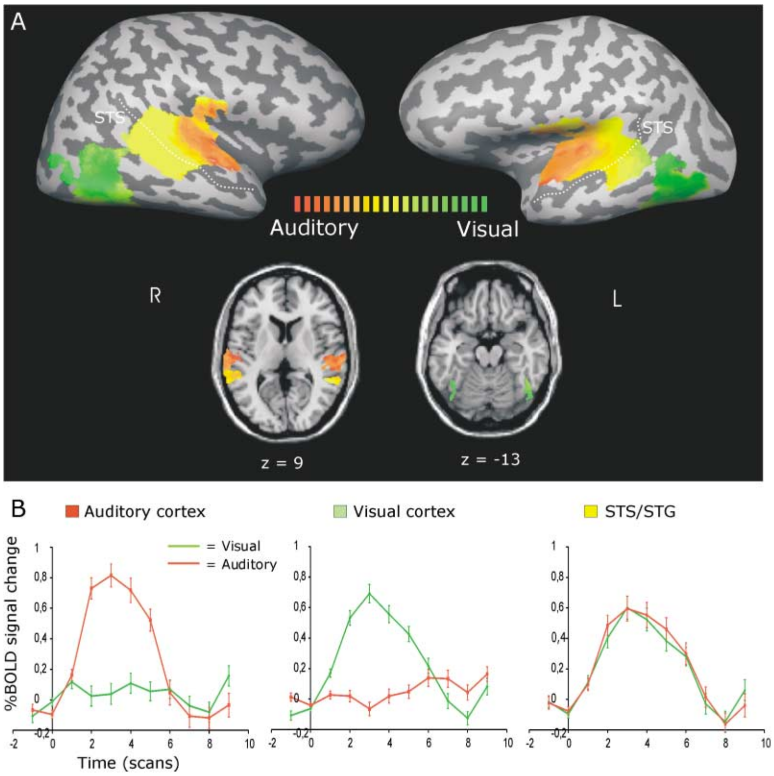

Figure 1. Group Results for the Unimodal Conditions

(A) Multisubject ($n = 16$) general linear model map of the unimodal predictors against baseline performed on the cortex-based aligned functional data. The statistical map is superimposed on the inflated cortical sheet of the individual brain used as target brain for the cortex-based intersubject alignment and on two axial slices of a template brain normalized in Talairach space. Color coding of the significantly activated voxels indicates relative contribution of auditory (red) and visual (green) unimodal stimulation to the signal change in each voxel. Similar contribution of both unimodal predictors is indicated by yellow.

(B) Averaged time courses of the BOLD response (in percent signal change) during unimodal auditory stimulation (red lines) and unimodal visual stimulation (green lines) in representative regions in auditory cortex, visual cortex, and superior temporal sulcus (STS). Error bars indicate SEM.

Group Results—Bimodal Conditions

The first step in identifying integration sites was to find regions that responded stronger to bimodal stimulation than to each unimodal condition, by performing the conjunction analysis of (bimodal congruent $>$ unimodal auditory) $\cap$ (bimodal congruent $>$ unimodal visual). This conjunction analysis will be referred to as *C2*. The second step was to constrain this analysis by including the requirement that integration sites should also respond significantly to each modality in isolation. We identified such regions with a conjunction analysis of the following contrasts: (bimodal congruent $>$ unimodal auditory) $\cap$ (bimodal congruent $>$ unimodal visual) $\cap$ (unimodal auditory $>$ baseline) $\cap$ (unimodal visual $>$ baseline). This conjunction analysis will be referred to as *C4*. To explore the effects of the relatedness of letters and speech sounds, we examined the contrast of bimodal congruent versus bimodal incongruent. This contrast analysis will be referred to as *congruency contrast*.

Figure 2 shows the results of the multisubject random-effect analyses of the congruency contrast [$t(15) = 3.75$, $q(FDR) < 0.05$] and the C2 and C4 analyses (shown at the same t value). As is shown in Figures 2A and 2C (green activation maps), the C4 analysis identified three superior temporal brain areas in the left hemisphere and one in the right hemisphere. The coordinates of corresponding regions in Talairach space are listed in Table 1. The averaged BOLD response time course (Figure 2B, STS/STG) indicates that the response to bimodal stimulation was stronger than to auditory and visual stimulation, and the response to both unimodal conditions stronger than baseline. It should be noted that the activation maps of the C2 and C4 analyses completely overlap (green regions), except for three regions in the anterior temporal cortex (shown in yellow): the planum polare (PP) bilaterally (left [−40, −3, −2] and right [47, 4, 0]) and left anterior STS (−49, −11, −7) are significant only in the C2 analysis. As is shown in the time course of the BOLD response (Figure 2B, PP/aSTS), only the response to congruent bimodal stimulation is significant in these anterior temporal areas.

The congruency contrast also revealed superior temporal regions in both hemispheres (Figures 2A and 2D). However, these regions were located more superiorly and anteriorly on the STG, on Heschl's sulcus (HS) and posterior from Heschl's sulcus on the planum temporale (PT) (Duvernoy, 1999; Kim et al., 2000). The BOLD response time courses of these regions (Figure 2B, PT/HS) reveal a highly interesting response pattern. The response to congruent letter-speech sound pairs was stronger than the response to speech sounds alone (ROI-based analyses: $t = 2.9$, $p < 0.005$ [right]; $t = 2.6$, $p < 0.01$ [left]), while the response to incongruent letter-speech sound pairs was reduced relative to the response to speech sounds alone (ROI-based analyses: $t = 5.1$, $p < 0.000$ [right]; $t = 5.4$, $p < 0.000$ [left]). Furthermore, there is only a very weak response to unimodal visual letters. These observations indicate that the response to speech sounds in early auditory areas (PT/HS) was modulated by the simultaneous visual presentation of letters, while letters presented in isolation did not activate these areas. No cross-modal modulation effects were observed in the visual cortex.

Individual Results—Interaction Effects

We selected four ROIs in each subject (if present): left ($n = 15$) and right ($n = 12$) PT/HS (based on the congruency contrast) and left ($n = 11$) and right ($n = 9$) STS/

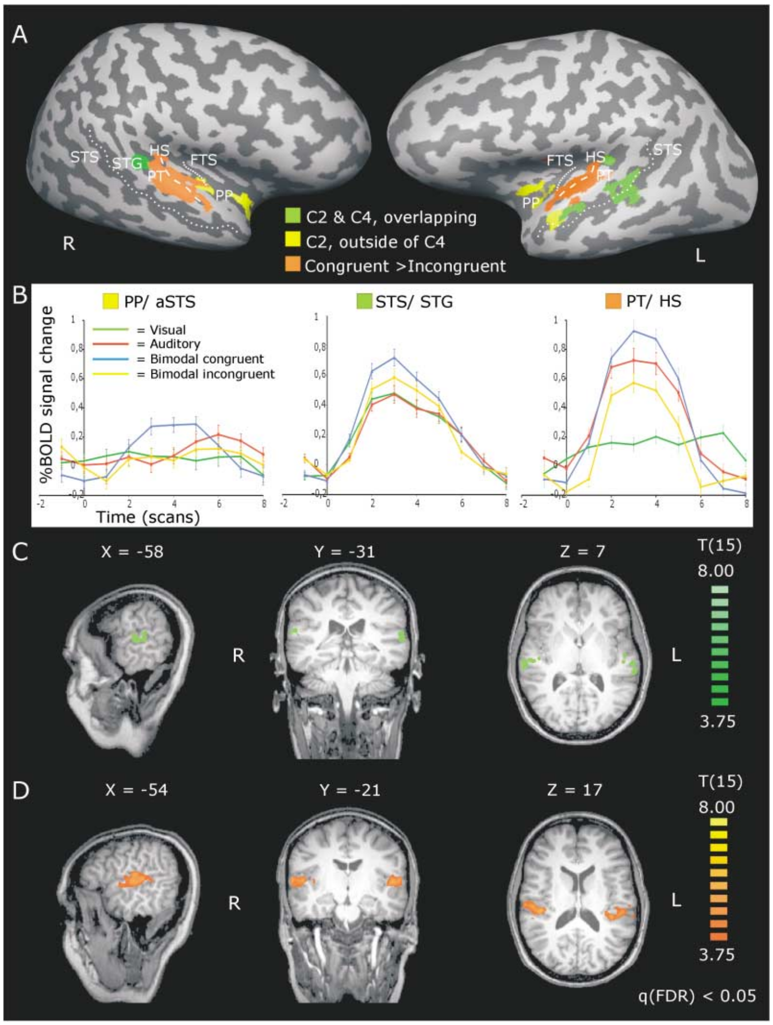

Figure 2. Group Results for the Conjunction and Congruency Analyses

(A) Multisubject ($n = 16$) general linear model maps of the two conjunction analyses (shown in yellow and green) and congruency contrast (shown in orange) performed on the cortex-based aligned functional data. The C2 and C4 maps completely overlap in the green regions; yellow regions are additionally activated in the C2 analysis. Maps are superimposed on the inflated cortical sheet of the individual brain used as target brain for the cortex-based intersubject alignment. The corresponding Talairach coordinates are listed in Table 1.

(B) Averaged time courses of the BOLD response (in percent signal change) during bimodal (congruent, blue lines; incongruent: yellow lines) and unimodal conditions (auditory, red lines; visual, green lines) in regions representative for the different maps shown in (A).

(C) Multisubject ($n = 16$) random-effect general linear model map of the C4 analysis superimposed on sagittal, coronal, and axial views of the Talairach normalized anatomical images of the target brain.

(D) Multisubject ($n = 16$) random-effect general linear model map of the congruent > incongruent analysis superimposed on sagittal, coronal, and axial views of the Talairach normalized anatomical images of the target brain. [C2, (bimodal congruent > unimodal auditory) $\cap$ (bimodal congruent > unimodal visual); C4, (bimodal congruent > unimodal auditory) $\cap$ (bimodal congruent > unimodal visual) $\cap$ (unimodal auditory > baseline) $\cap$ (unimodal visual > baseline); PP, planum polare; (a)STS, (anterior) superior temporal sulcus; STG, superior temporal gyrus; PT, planum temporale; HS, Heschl's sulcus; FTS, first transverse temporal sulcus].

STG (based on the C4 analysis). Talairach coordinates and significance levels for all selected ROIs in individual subjects are listed in Table 2. We estimated b values for all four predictors by individual ROI-based GLM analyses (b_a, unimodal auditory; b_v, unimodal visual; b_{avi}, bimodal incongruent; b_{avc}, bimodal congruent). Figure 3 shows the normalized b values ($b_a + b_v = 1$) averaged over subjects for all four predictors in the four ROIs.

Table 1. Talairach Coordinates of Superior Temporal Integration Sites, Revealed by the Conjunction Analyses, and Comparable Regions Reported by Other Studies

Study	Stimuli	Region	x	y	z
a	letters and speech sounds	left STS	−54	−48	9
		left STS	−46	−19	2
		left STG	−43	−43	23
		right STG	52	−33	18
b	letters and speech sounds	left STS	−53	−31	0
		right STS	48	−31	6
c	audiovisual speech	left STS	−49	−50	9
d	audiovisual speech	left STS	−56	−49	9
e	complex objects	left STS/MTG	−50	−55	7
f	white noise/ checkerboard	left STS	−51	−36	9

a, present study; b, Raij et al. (2000); c, Calvert et al. (2000); d, Sekiyama et al. (2003); e, Beauchamp et al. (2004); f, Calvert et al., 2001.

Superadditivity [$b_{av} > (b_a + b_v)$] was observed for the congruent bimodal condition in the PT/HS bilaterally [left: $t(14) = 2.5$, $p = 0{,}025$; right, $t(11) = 2.9$, $p = 0.014$] but not in STS/STG, where b_{avc} is significantly less than ($b_a + b_v$) in the right hemisphere [left: $t(10) = -1.7$, $p = 0.13$; right, $t(8) = -3.3$, $p = 0.011$]. However, in the STS/STG, b_{avc} was significantly higher than b_a [left: $t(10) = 10.1$, $p = 0{,}000$; right $t(8) = 9.9$, $p = 0.000$] and b_v [left: $t(10) = 14.7$, $p = 0.000$; right: $t(8) = 14.2$, $p = 0.000$]. Response suppression ($b_{av} < [\max (b_a, b_v)]$), was observed for b_{avi} in the PT/HS [left: $t(14) = 7.1$, $p = 0.000$; right: $t(11) = 9.0$, $p = 0.000$] but not in the STS/STG [left: $t(10) = -1.3$, $p = 0.24$; right: $t(8) = -1.9$, $p = 0.1$]. In the STS/STG, b_{avi} was subadditive [$b_{av} < (b_a + b_v)$] [left: $t(10) = -12.9$, $p = 0.000$; right: $t(8) = -7.4$, $p = 0.000$]. Figure 3 furthermore reveals a difference in relative proportion of the b values of both unimodal conditions in PT/HS and STS/STG. In STS/STG, b values of both unimodal conditions were equally high, while in PT/HS, the b value of the unimodal visual condition was very low relative to the b value of the unimodal auditory

Table 2. Talairach Coordinates and Significance Levels for ROIs in Individual Subjects

ROIs in Planum Temporale/Heschl's Sulcus

	Left Hemisphere					Right Hemisphere				
Subject	x	y	z	t[a]	p[a]	x	y	z	t[a]	p[a]
MP	−59	−23	16	2.6	0.005	52	−14	13	2.8	0.01
HM	−59	−24	12	3.9	0.001	–	–	–	–	–
SO	−58	−17	10	3.2	0.01	61	−16	13	3.1	0.005
PP	−61	−22	9	4.2	0.0000	60	−14	7	4.8	0.000
SS	−49	−11	7	2.8	0.01	52	−16	11	2.7	0.01
MJ	−47	−20	7	2.6	0.01	–	–	–	–	–
MA	–	–	–	–	–	–	–	–	–	–
JP	−32	−30	9	3.5	0.001	39	−20	9	3.7	0.0005
BM	−47	−29	16	2.0	0.05	64	−11	14	3.4	0.001
CB	−49	−11	10	2.0	0.05	–	–	–	–	–
MS	−47	−16	5	3.6	0.0005	51	−16	12	4.3	0.0000
MH	−52	−31	17	3.7	0.0005	37	−22	16	2.4	0.05
NH	−57	−25	7	3.1	0.005	48	−17	15	2.4	0.05
KP	−56	−19	10	2.7	0.01	64	−19	8	2.2	0.05
JK	−58	−26	3	4.0	0.0001	59	−12	17	3.6	0.0005
RW	−51	−12	4	3.6	0.0005	39	−13	5	3.9	0.0005
Average	−51	−24	10			52	−16	12		
(Standard deviation)	(8)	(9)	(5)			(10)	(3)	(4)		

ROIs in Superior Temporal Sulcus/Gyrus

	Left Hemisphere							Right Hemisphere						
Subject	x	y	z	t[b]	p[b]	t[c]	p[c]	x	y	z	t[b]	p[b]	t[c]	p[c]
MP	−53	−50	15	2.1	0.05	2.0	0.05	–	–	–	–	–	–	–
HM	−56	−41	6	2.6	0.05	2.2	0.05	66	−22	0	2.7	0.01	4.0	0.0001
SO	–	–	–	–	–	–	–	47	−31	7	2.1	0.05	2.3	0.05
PP	−53	−44	10	2.7	0.01	3.5	0.001	53	−34	9	2.1	0.05	2.8	0.01
SS	−60	−28	−2	2.4	0.05	2.9	0.005	–	–	–	–	–	–	–
MJ	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MA	–	–	–	–	–	–	–	–	–	–	–	–	–	–
JP	−52	−23	2	2.2	0.05	3.7	0.0005	–	–	–	–	–	–	–
BM	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CB	−57	−22	10	2.4	0.05	2.0	0.05	56	−25	14	2.3	0.05	2.7	0.01
MS	−56	−41	6	2.3	0.05	5.6	0.0000	58	−36	18	3.3	0.005	5.2	0.0000
MH	−53	−32	17	2.3	0.05	3.2	0.005	52	−24	21	2.0	0.05	2.4	0.05
NH	–	–	–	–	–	–	–	62	−37	14	2.4	0.05	2.4	0.05
KP	−46	−16	1	2.4	0.05	2.1	0.05	–	–	–	–	–	–	–
JK	−59	−39	4	2.4	0.05	3.6	0.0005	54	−25	7	2.5	0.05	3.3	0.005
RW	−54	−22	3	2.5	0.05	3.8	0.0005	39	−42	18	2.2	0.05	2.3	0.05
Average	−54	−33	7					54	−31	12				
(Standard deviation)	(4)	(11)	(6)					(8)	(7)	(7)				

[a] t and p values for the Congruent versus Incongruent contrast.
[b] t and p values for the Congruent versus Auditory contrast.
[c] t and p values for the Congruent versus Visual contrast.

condition. It should be noted that the lack of superadditivity in the STS/STG might be due to the strong response to the isolated unimodal conditions (see Discussion).

Individual Results—Detailed Localization of Modulation Effect

In 15 of the 16 subjects, a region was observed in the left hemisphere PT/HS that showed modulation of the response to speech sounds depending on the congruency of the letters and speech sounds, but no response to unimodally presented letters. Twelve of these 15 subjects showed a similar region in the right hemisphere (see Table 2). Figure 4 shows the location of the ROIs in PT and HS in six representative individual subjects (dark blue regions), together with the whole region activated by speech sounds (light blue regions). An interesting observation is that only a subregion of the area activated by speech sounds was modulated by visual letter information. In both hemispheres, the location of regions showing this modulation effect was highly consistent across all subjects, on HS extending on the PT. This location is just posterior and lateral to the primary auditory cortex, which is located on HG, between the first transverse sulcus (FTS) and Heschl's sulcus (Formisano et al., 2003), and may correspond to what has been described as lateral belt cortex in the macaque monkey (Kaas and Hackett, 2000).

Discussion

The aim of the present study was to investigate the neural correlates of the integration of letters and speech sounds. The most interesting finding is a cross-modal

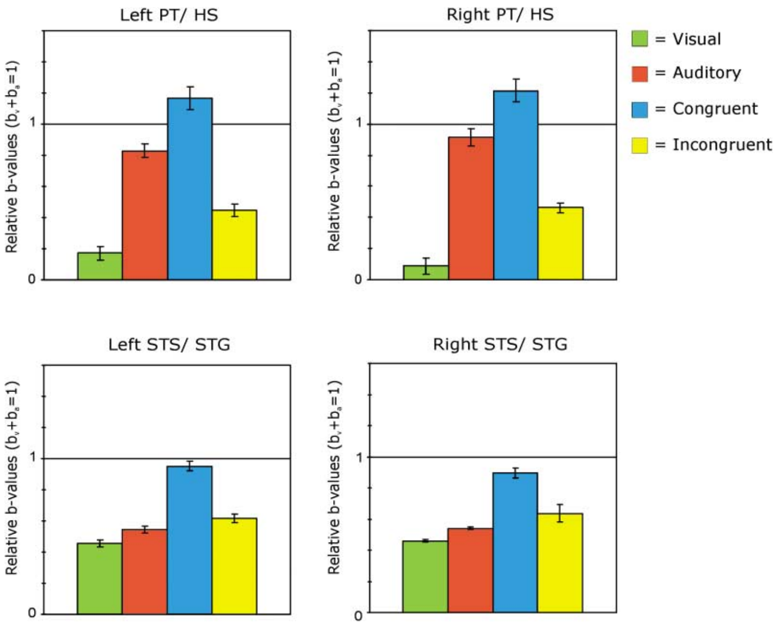

Figure 3. Results of the Interaction Analysis Performed on b Values Estimated by Individual ROI-Based GLM Analyses

Normalized b values for all four predictors ($b_v + b_a = 1$) are shown for the ROIs in the planum temporale (PT)/Heschl's sulcus (HS) and superior temporal sulcus (STS)/superior temporal gyrus (STG). Left PT/HS, n = 15; right PT/HS, n = 12; left STS/STG, n = 11; right STS/STG, n = 9. Error bars indicate SEM across subjects.

modulation of the response to speech sounds in modality-specific early auditory cortex (HS and PT) by simultaneous visual presentation of congruent or incongruent letters. Integration sites were identified in the STG and STS. These regions showed a heteromodal response pattern, i.e., responded significantly to both unimodal conditions and in addition responded stronger to bimodal than to unimodal letters and speech sounds. Furthermore, anterior temporal regions were revealed that only responded to bimodal congruent letters. A schematic summary of our findings regarding the integration of letters and speech sounds is depicted in Figure 5.

Modality-specific activation to letters and speech sounds was found in corresponding sensory processing areas. The inferior occipital-temporal cortex (green regions, Figure 1) responded bilaterally to unimodally presented letters but not to unimodally presented speech sounds. This location is consistent with other fMRI studies that investigated single letter processing (Gauthier et al., 2000; Gros et al., 2001; Longcamp et al., 2003; Polk et al., 2002). Modality-specific auditory activation was found in superior temporal cortex (Heschl's gyrus, planum temporale, and the middle part of the STG, see red regions in Figure 1). These regions were activated by speech sounds but not by letters. The location of these regions is consistent with areas involved in speech sound processing reported by other imaging studies (Demonet et al., 1992; Jäncke et al., 2002; Suzuki et al., 2002; Tervaniemi et al., 2000).

Bilateral regions in the STS and posterior STG responded to the unimodal presentation of both letters and speech sounds. Interestingly, these regions are anatomically located between the visual and auditory processing regions where modality-specific responses were observed (yellow regions, Figure 1). These observations indicate that multimodal convergence between letter and speech sound processing occurs in the STS/STG. This is consistent with the overlap between activations for the perception of speech and lip movements in STS and STG reported by Wright and colleagues (Wright et al., 2003). Because large numbers of neurons are sampled simultaneously with fMRI, responsiveness of a region to stimulation in different modalities may reflect responses of subpopulations of modality-specific neurons within the same voxels instead of neuronal convergence (King and Calvert, 2001; Meredith, 2002). However, single-cell recordings in primates have shown that neurons in STS and STG receive inputs from different sensory modalities (Baylis et al., 1987; Bruce et al., 1981; Hikosaka et al., 1988; Schroeder and Foxe, 2002). Taking together our finding of converging activations and the physiological evidence from primate single-cell recordings, it seems plausible that activations evoked by letters and speech sounds converge at the neuronal level in the STS/STG.

Following the assumption that anatomical convergence of multisensory input is a prerequisite for integration (Meredith, 2002; Raij et al., 2000), integration sites should respond to each modality in isolation (to determine convergence) and stronger to bimodal than to unimodal stimulation (to determine integration). Several regions were revealed in superior temporal cortex bilaterally (located on STS and STG, see green regions in Figure 2) that responded in this way, providing candidates for integration sites of letters and speech sounds. The location of these superior temporal regions is remarkably similar to those involved in the integration of audiovisual speech (Calvert et al., 2000; Sekiyama et al., 2003; Wright et al., 2003), audiovisual information about complex objects (Beauchamp et al., 2004), and temporally matched meaningless audiovisual stimuli (Calvert et al., 2001) (see Table 1). Our results therefore support the general role of the STS/STG in integrating audiovisual identity information (Beauchamp et al., 2004; Calvert, 2001).

Our analyses also revealed areas which were active during bimodal stimulation but neither during visual nor auditory unimodal stimulation. Brain areas with such a

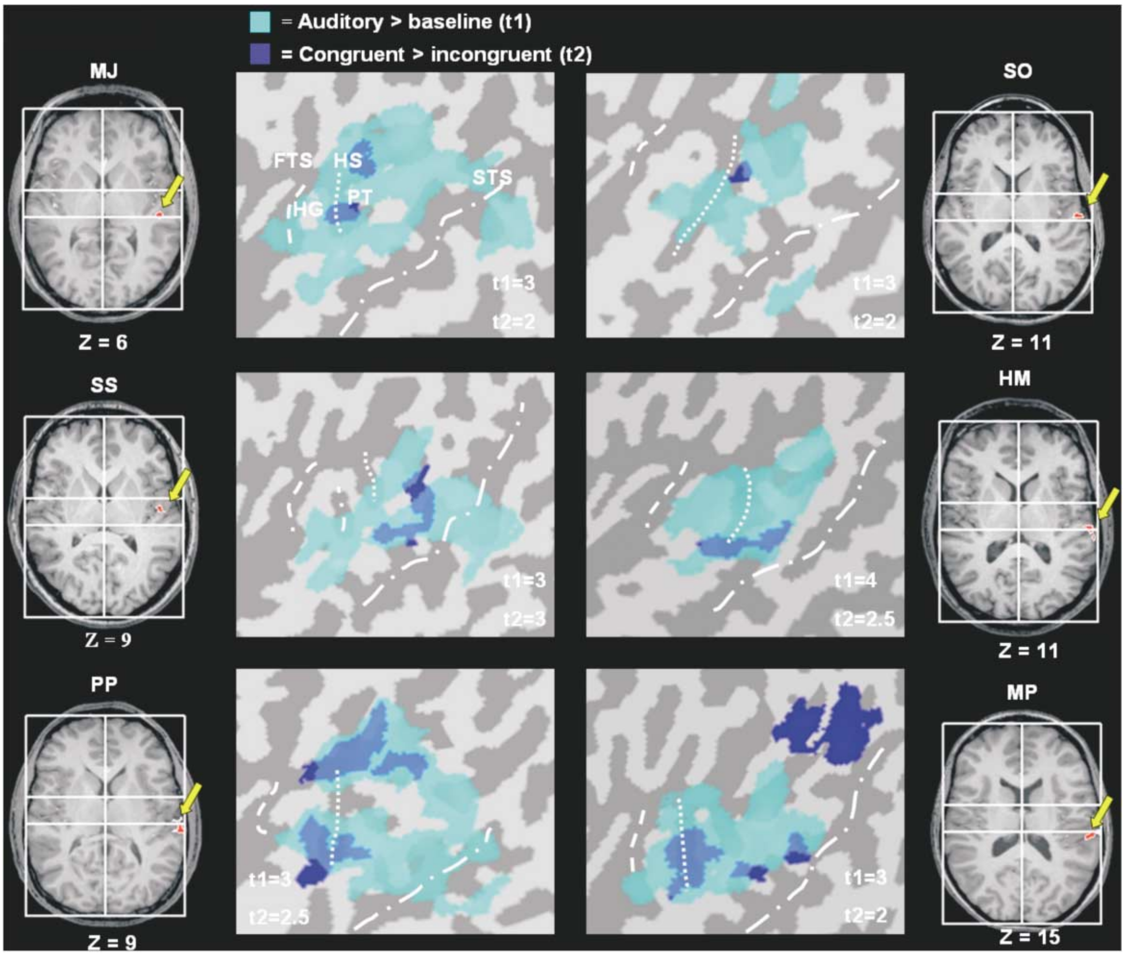

Figure 4. Location of ROIs in Early Auditory Cortex in the Left Hemisphere in Six Representative Individual Subjects

Statistical maps for unimodal auditory > baseline (light blue) and bimodal congruent > bimodal incongruent (dark blue) are superimposed on the inflated cortical sheets of the individual subjects. The congruency contrast is also superimposed on axial views of the individual Talairach normalized anatomical images (orange clusters indicated by yellow arrows). t values for both contrasts are given in the figure. STS, superior temporal sulcus; PT, planum temporale; HS, Heschl's sulcus; FTS, first transverse temporal sulcus.

response profile would indicate that they contain cells exhibiting conjunctive coding (Hinton et al., 1986). Without the constraint that modality-specific responses should be significant, we indeed revealed additional activation clusters, located in the anterior temporal cortex (see Figure 2A, yellow regions). Interestingly, these regions responded only during congruent but not during incongruent bimodal stimulation (see Figure 2B). This unexpected selectivity for congruent bimodal stimulation might indicate that these regions signal the successful binding of cross-modal information into a unified percept. Future work is necessary to investigate the detailed properties of these regions. It would, for example, be interesting to know whether some or all subregions are specifically involved in coding congruent letter-sound stimulation or whether they would also respond during other congruent, but not incongruent, visual and auditory information. Furthermore, it would be interesting to find out whether these anterior temporal regions receive input directly from the modality-specific processing regions, which is expected if conjunctive coding indeed takes place, or whether the selective response to congruent information is determined by input from heteromodal STS/STG.

To explore the effects of relatedness between the stimuli in both modalities, we examined the contrast of congruent versus incongruent bimodal stimulation. This comparison revealed cross-modal modulation effects in early auditory cortex (HS and PT), expressed by enhancement and suppression of the response to speech sounds depending on the congruency of the simultaneously presented letters. Congruent combinations of letters and speech sounds elicited a response that significantly exceeded the response to speech sounds alone, while the response to incongruent combinations of letters and speech sounds was significantly weaker than the response to speech sounds alone. Individual analyses indicated a highly consistent localization of this cross-modal modulation effect, just posterior and lateral to the primary auditory cortex (see Figure 4 and Table 2). This location indicates that visual letter information can influence auditory processing of speech sounds at a very early stage. Another remarkable observation is that these regions did not respond to letters alone, sug-

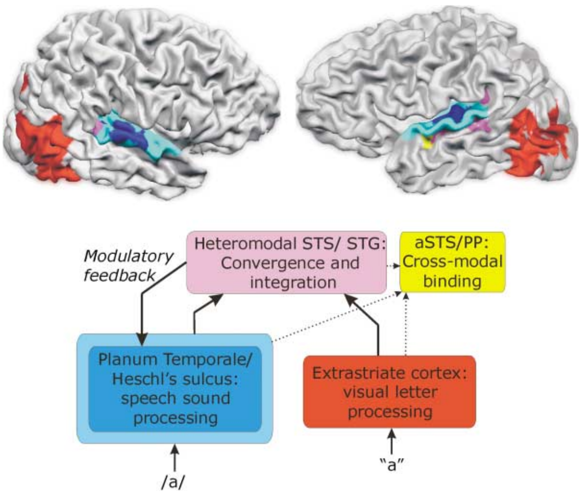
Modulatory feedback
Heteromodal STS/ STG: Convergence and integration
aSTS/PP: Cross-modal binding
Planum Temporale/ Heschl's sulcus: speech sound processing
Extrastriate cortex: visual letter processing
/a/
"a"

Figure 5. Schematic Summary of Our Findings on the Integration of Letters and Speech Sounds

Multisubject statistical maps are superimposed on the folded cortical sheet of the individual brain used as target brain for the cortex-based intersubject alignment: unimodal visual > unimodal auditory (red), unimodal auditory > unimodal visual (light blue), congruent > incongruent (dark blue), and the conjunction of bimodal > unimodal and unimodal > baseline (C4, violet). A schematic description of our findings and interpretation based on the functional neuroanatomical data is shown in the lower panel. Dashed lines indicate that aSTS/PP may either receive input directly from the modality-specific regions or from the heteromodal STS/STG. (a)STS, (anterior) superior temporal sulcus; STG, superior temporal gyrus; PP, planum polare; /a/, speech sound *a*; "a," letter *a*.

gesting that the influence of letters on the processing of speech sounds in early auditory cortex is indirect and may reflect an outcome of prior integration in heteromodal STS/STG rather than integration itself.

It should be noted that multisensory integration at the neuronal level is usually inferred on the basis of an interaction test [$AV \neq (A + V)$] (see e.g., Wallace et al., 1996). A similar method has recently been used to detect integration from EEG (e.g., Fort et al., 2002), MEG (e.g., Raij et al., 2000), and fMRI data (Calvert et al., 2000, 2001; Wright et al., 2003). Following this criterion, positive interaction is determined by superadditivity [$AV > (A + V)$] and negative interaction by subadditivity [$AV < (A + V)$] or response suppression ($AV <$ [max (A, V)]), where [max (A, V)] indicates the largest of the unimodal responses Calvert et al., 2001). We tested for interaction effects using the b values for the four predictors estimated by single-subject GLM analyses in selected ROIs in STS/STG and PT/HS (Figure 3). Our results show that in left and right PT/HS, the b value of the congruent bimodal predictor is superadditive and the b value of the incongruent bimodal predictor shows response suppression. In the left and right STS/STG, the b value of the congruent bimodal predictor is significantly higher than either unimodal b value but does not reach superadditivity. Furthermore, the b value of the incongruent bimodal predictor is subadditive but does not show response suppression.

However, results from interaction analyses performed on fMRI data should be interpreted with caution (Beauchamp et al., 2004; Calvert, 2001; King and Calvert, 2001; Wright et al., 2003). Because of the intrinsic nature of the BOLD response and its limited dynamic range, it is possible that a superadditive change in the response at the neuronal level is not reflected in a similarly superadditive change in the BOLD response. Therefore, superadditivity is more likely to be found when the response to one or both of the unimodal conditions is weak or absent (or even negative) than when the BOLD response to both modalities is separately already strong. This prediction is confirmed by the findings of Wright et al. (2003) and by the different response patterns we find in the anterior temporal cortex, PT/HS, and the STS/STG (see Figures 2B and 3). In the PT/HS, the very weak response to unimodal visual stimuli leaves opportunity for congruent bimodal stimuli to evoke a superadditive BOLD response. In the anterior temporal cortex, neither unimodal condition evokes a significant response, so the response evoked by the congruent bimodal condition exceeds the sum of the unimodal responses even more extremely. Conversely, the response to both unimodal conditions is considerable in the STS/STG, preventing the increased response to bimodal congruent stimulation from reaching superadditivity. These observations clearly indicate the importance of also inspecting the unimodal responses when interpreting interaction effects. Because of the difficulties in interpreting interactions measured with fMRI, we decided to use the conjunction analyses to determine cross-modal integration at a map level. A similar approach for inferring cross-modal integration from fMRI data has been used by another recent fMRI study (Beauchamp et al., 2004). Based on the combination of our conjunction and congruency analyses we suggest that integration of letters and speech sounds is performed in heteromodal regions in STS/STG and that the outcome of this integration process is projected back to the putative "modality-specific" auditory regions in HS and PT and modulates the processing of speech sounds.

The idea of interplay between modality-specific and heteromodal brain areas during cross-modal integration is supported by other fMRI findings (Bushara et al., 2003; Macaluso et al., 2000; Calvert et al., 1999, 2000). The cross-modal modulation effect we observe in phoneme-processing regions in the auditory cortex is consistent with the suggestion that integration sites in STS/STG modulate processing in sensory cortices by feedback projections during audiovisual speech perception (Calvert et al., 1999, 2000). We think that the enhancement (suppression) of activity that is observed in these regions in the case of congruent (incongruent) visual presentation of letters is a consequence of the presence of language-related audiovisual functional connections developed during reading acquisition and daily used. Whether and where in the brain similar effects are present for other audiovisual pairs is an interesting empirical question that deserves to be investigated in future studies. In sum, our findings indicate that the integration of letters and speech sounds is performed by a neural mechanism similar to the mechanism for the integration of speech with lip movements (or possibly an even more general audiovisual integration system). This similarity is crucial and profound, since it supports the hypothesis that the efficient use of letter-sound associations is possible because the brain can make use of a naturally evolved mechanism for associating visual information with speech.

A feedback mechanism implies that the modulation effect observed in HS and PT should occur later in time than the convergence and integration in STS/STG. Unfortunately, fMRI does not provide temporal information that is accurate enough to verify this implication. However, the time course information on audiovisual letter integration provided by Raij and colleagues (Raij et al., 2000) does support a feedback mechanism: convergence and interaction effects started at 225 ms, while differences in interaction effects for matching and nonmatching letters and speech sounds did not occur until 380–450 ms. More direct support for our proposed feedback mechanism comes from a recent report by Schroeder and Foxe (2002). They investigated the laminar profile and multisensory response properties of neurons in the posterior auditory association cortex (presumably corresponding to our PT/HS regions) and the superior temporal sulcus (presumably corresponding to our STS/STG regions) of the macaque monkey. In auditory association cortex, a laminar input profile for visual stimulation was observed, indicating feedback projections. In the STS, laminar profiles indicated feedforward convergence of visual and auditory information. Consistent information was provided by the response latencies of visual and auditory input: responses to visual and auditory stimulation in the STS were coincident, while visual responses were delayed relative to auditory responses in the auditory association cortex.

The modulation effect observed in the auditory association cortex was not observed in the visual association cortex. This implies that the outcome of the integration of letters and speech sounds is projected back to influence selectively only early auditory processing levels, but not lower-level visual processing. In contrast, both sensory systems seem to be related reciprocally during audiovisual speech perception (Calvert et al., 1999). An asymmetry in the representations of associations between letters and speech sounds has been reported at the behavioral level before (Dijkstra et al., 1993; Hardy et al., 1972). A possible explanation for the observed unidirectional influence is that speech sounds are continuous in time and more variable and therefore more difficult to recognize than discrete and invariable letter symbols (Liberman, 1992). This is consistent with the finding of Sekiyama et al. (2003) that visual information exerts a stronger influence on auditory processing when speech is less intelligible. The direction of modulatory effects between letters and speech sounds may depend on the temporal synchrony of the stimuli (Dijkstra et al., 1993; Jones and Callan, 2003), a possibility that will need further investigation.

A remark should be made about the possible involvement of frontal and parietal brain regions in the integration of letters and speech sounds. Frontal (Broca's area, premotor cortex, anterior cingulate cortex) and parietal (inferior and posterior parietal cortex) activation has been reported during lip reading and audiovisual speech perception (Calvert et al., 1997, 2000; Calvert and Campbell, 2003; Jones and Callan, 2003; Olson et al., 2002; Paulesu et al., 2003; Sekiyama et al., 2003) and are associated with speech sound processing and attention. Because of methodological considerations (see Experimental Procedures), our functional scans did not cover the whole brain. Therefore, we cannot exclude the possibility that frontal and parietal brain regions may also play a role in the proposed mechanism for the integration of letters and speech sounds.

Conclusion

By using an fMRI design that allowed the investigation of processing in auditory, visual, and heteromodal temporal brain regions with high spatial accuracy, we revealed a functional neuronal mechanism for the integration of letters and speech sounds. Modality-specific processing was observed in superior temporal and occipital temporal cortices, convergence and integration in the heteromodal STS/STG. Furthermore, we revealed anterior temporal regions that exclusively responded to bimodal congruent letters. A most interesting finding was that subregions of early auditory cortex involved in speech sound processing, located on Heschl's sulcus and the planum temporale, were influenced by the congruency of simultaneously presented letters and speech sounds. Because these regions did not respond to letters alone, we interpret this influence as feedback modulation from heteromodal STS/STG regions where integration took place. Our data show that the integration of culturally defined associations between letters and speech sounds recruits a neural mechanism similar to the naturally evolved neural mechanism for the integration of speech information with lip movements.

Experimental Procedures

Subjects

Sixteen healthy native Dutch subjects (3 male, mean age 22 ± 2.4, range 19–27) participated in the present study. All subjects were university students enrolled in an undergraduate study program. We selected subjects based on a questionnaire including questions concerning (present or history of) 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.

Stimuli and Task Design

Stimuli were speech sounds corresponding to single letters and visually presented single letters. Speech sounds were digitally recorded (sampling rate 44.1 kHz, 16 bit quantization) from a female speaker. Recordings were band-pass filtered (180–10,000 Hz) and resampled at 22.05 kHz. Only speech sounds that were recognized 100% correctly in a pilot experiment ($n = 10$) were selected for the fMRI experiment. Selected consonants were b, d, g, h, k, l, n, p, r, s, t, z. Selected vowels were a, e, i, y, o, u. Average duration of the speech sounds was 352 (±5) ms, the average sound intensity level was 71.3 (±0.2) dB. Visual stimuli were lower-case letters corresponding to the 18 speech sounds. White letters were presented for 350 ms on a black background in the center of a computer screen, printed in "Arial" font at letter-size 40. During fixation periods and scanning, a white fixation cross was presented in the center of the screen.

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 s) were composed of four miniblocks of 5 s (see Figure 6). In the first 1.1 s of each miniblock, one brain volume was acquired (see "Scanning Procedure"). 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 3.9 s, five stimuli were presented with a stimulus onset asynchrony (SOA) of 700 ms.

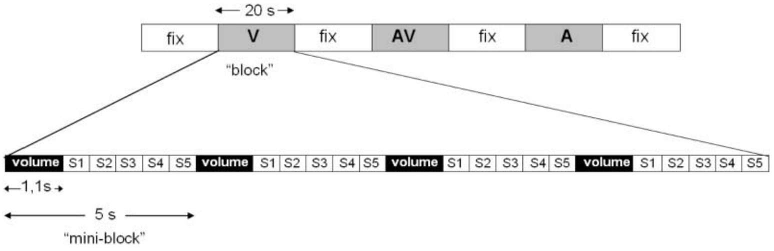

Figure 6. Schematic Description of the Experimental Design

Experimental blocks of 20 s were composed of four miniblocks of 5 s. Each miniblock starts with the acquisition of one volume scan followed by five stimuli presented in silence. Experimental blocks were alternated with fixation periods of 20 s. fix, fixation period; V, visual experimental block; AV, audiovisual experimental block; A, auditory experimental block; S, stimulus.

This stimulus presentation design was adapted from Jäncke et al. (2002). Stimulus presentation was synchronized with the scanner pulses using the software package "Presentation" (http://neurobehavioralsystems.com).

Vowels and consonants were presented in separate blocks. Every experimental condition was repeated twice in each run. Experimental blocks were alternated with fixation periods of 20 s; every run started and ended with a fixation period. Two runs (80 trials per condition) were acquired in 7/16 subjects; four runs (160 trials per condition) were acquired in 9/16 subjects. The order of the experimental blocks was pseudorandomized within runs and counterbalanced across runs.

Subjects passively listened to and/or viewed the stimuli. Since we were interested in fundamental brain responses to the processing and integration of letters and speech sounds, a passive perception design was chosen because it avoids interaction between activity related to stimulus processing and task-related activity due to cognitive factors (Calvert et al., 2000).

Scanning Procedure

Imaging was performed on a 3 Tesla whole-body system (Magnetom Trio, Siemens Medical Systems, Erlangen, Germany). In each subject, two or four runs of 71 volumes were acquired using a BOLD-sensitive EPI sequence (matrix, $64 \times 64 \times 18$; voxel size, $3 \times 3 \times 4$ mm^3, FOV, 192 mm^2; TE/TRslice, 32/63 ms; FA, 75°). Sequence scanning time was 1.1 s, interscan gap was 3.9 s, resulting in a TR (sequence repeat time) of 5 s. A slab of 18 axial slices (slab thickness 7.2 cm) was positioned in each individual such that the entire visual and auditory cortices were covered, based on anatomical information from a scout image of seven sagitally oriented slices. In most of the subjects, the slab did not cover the whole brain, excluding the superior frontal and parietal cortices. We reduced the number of slices to maximize scanning speed and silent delay time for stimulus presentation without increasing excessively the volume TR. A high-resolution structural scan (voxel size $1 \times 1 \times 1$ mm^3) 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). Clustering of the volume acquisition in the beginning of each TR provides a period of silence (3.9 s in our experiment) between successive scans in which stimulus perception is uncontaminated by EPI noise. This stimulation procedure is demonstrated to be highly efficient for studying the auditory cortex with fMRI (Jäncke et al., 2002; Shah et al., 2000).

Image Analysis and Visualization

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 fewer 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 were transformed into Talairach space.

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, inflated, and flattened. 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. While functional areas do not precisely follow cortical landmarks, it has been shown that a cortical matching approach substantially improves statistical analysis across subjects by reducing anatomical variability (Fischl et al., 1999). Following Fischl et al. (1999), we morphed the reconstructed folded cortical representations of each subject and hemisphere into a spherical representation, which provides a parameterizable surface well suited for across-subject nonrigid alignment. Each vertex on the sphere (spherical coordinate system) corresponds to a vertex of the folded cortex (Cartesian coordinate system) and vice versa. The curvature information computed in the folded representation is preserved as a curvature map on the spherical representation. The curvature information (folding pattern) is smoothed along the surface to provide spatially extended gradient information driving intercortex alignment minimizing the mean squared differences between the curvature of a source and a target sphere. The alignment proceeds iteratively following a coarse-to-fine matching strategy, which starts with highly smoothed curvature maps and progresses to only slightly smoothed representations. Starting with a coarse alignment as provided by Talairach space, this method ensures that the smoothed curvature of the two cortices possess enough overlap for a locally operating gradient-descent procedure to converge without user intervention. Visual inspection and a measure of the averaged mean squared curvature difference revealed that the alignment of major gyri and sulci was achieved reliably by this method. Smaller structures, visible in the slightly smoothed curvature maps, were not completely aligned, reflecting idiosyncratic differences between the subjects' brains, such as continuous versus broken characteristics. The established correspondence mapping between vertices of the cortices was used to align the time courses for multisubject GLM data analysis.

Single-subject and multisubject analyses were performed by multiple linear regression of the BOLD-response time course in each voxel using four predictors: unimodal visual, unimodal auditory, bimodal congruent, and bimodal incongruent. Predictor time courses were adjusted for the hemodynamic response delay by convolution with a hemodynamic response function (Boynton et al., 1996) (delta 2.5, tau 1.25). Random-effect analyses were performed on the group data ($n = 16$), enabling generalization of the statistical inferences to the population level. Multisubject statistical maps were thresholded using the false discovery rate (FDR; Genovese et al., 2002). For ROI-based analyses, t statistics and corresponding p values are reported.

To reveal modality-specific responses, the unimodal visual and unimodal auditory predictors were contrasted against baseline. Results of the unimodal predictors against baseline were visualized using a two set relative contribution map (Figure 1). The whole colored region in these maps depicts significant explanation of the signal variance by the visual or auditory predictors against baseline; the color coding within this map indicates where the visual predictor contributes more to explaining the variance, where the auditory predictor contributes more, and where they both contribute equally. The relative contribution value in each voxel was computed by $(b_v - b_a)/(b_v + b_a)$.

To identify integration sites, we used conjunction analyses to extract conjoint activation in the bimodal conditions contrasted to each unimodal condition. At each voxel, a new statistical value was computed as the *minimum* of the statistical values obtained from all the specified contrasts. The first step in identifying integration sites was to find voxels that were significantly more active during audiovisual stimulation as compared to both modalities in isolation. For this purpose, we performed the conjunction analysis with the following two contrasts: (bimodal congruent $>$ unimodal auditory) $\cap$ (bimodal congruent $>$ unimodal visual). In the second step, we included the constraint that integration sites should show convergence of auditory and visual activity, i.e., respond to both unimodal conditions in isolation. In other words, two contrasts were added to the conjunction analysis: (bimodal congruent $>$ unimodal auditory) $\cap$ (bimodal congruent $>$ unimodal visual) $\cap$ (unimodal auditory $>$ baseline) $\cap$ (unimodal visual $>$ baseline).

To determine interaction effects, we estimated b values for all four predictors by individual ROI-based GLM analyses and contrasted the bimodal predictors to the sum of the unimodal b values ($b_a + b_v$). For this purpose, we normalized all b values by dividing them by ($b_a + b_v$). After this normalization, superadditivity is defined by $b_{av} > 1$, and subadditivity by $b_{av} < 1$, and response suppression by ($b_{av} < [\max (b_a, b_v)]$), where $[\max (b_a, b_v)]$ indicates the largest of the unimodal responses (Calvert et al., 2001). We tested for superadditivity and subadditivity by one-sample t tests (b_{av} against 1) and for response suppression by paired-samples t tests [b_{av} against max (b_a, b_v)].

To find out which brain regions respond differently to congruent and incongruent combinations of letters and speech sounds, the contrast of bimodal congruent versus bimodal incongruent was performed.

Acknowledgments

The authors are grateful to Holger Mitterer for his help with preparing the stimuli, to Paul Gaalman for his technical assistance, and to Peter Hagoort for providing access to the facilities of the F.C Donders Centre. The reported research was supported by grant 048.011.046 of the Dutch Organisation for Scientific Research (NWO) awarded to L.B.

Received: January 5, 2004
Revised: May 10, 2004
Accepted: June 16, 2004
Published: July 21, 2004

References

Baylis, G.C., Rolls, E.T., and Leonard, C.M. (1987). Functional subdivisions of the temporal lobe neocortex. J. Neurosci. *7*, 330–342.

Beauchamp, M., Lee, K., Argall, B., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron *41*, 809–823.

Blomert, L. (2002). Dyslexie: Stand van Zaken (Dyslexia: State of Affairs in The Netherlands). Report for the Dutch Ministry of Health. In Dyslexie naar een vergoedingsregeling, R. Reij, ed. (Amstelveen: Dutch Health Care Insurance Board).

Boynton, G.M., Engel, S.A., Glover, G.H., and Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. *16*, 4207–4241.

Bruce, C., Desimone, R., and Gross, C.G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J. Neurophysiol. *46*, 369–384.

Bushara, K.O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., and Hallett, M. (2003). Neural correlates of cross-modal binding. Nat. Neurosci. *6*, 190–195.

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

Calvert, G.A., and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. *15*, 57–71.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science *276*, 593–596.

Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iversen, S.D., and David, A.S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport *10*, 2619–2623.

Calvert, G.A., Campbell, R., and Brammer, M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. *10*, 649–657.

Calvert, G.A., Hansen, P.C., Iversen, S.D., and Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage *14*, 427–438.

Demonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., Rascol, A., and Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. Brain *115*, 1753–1768.

Dijkstra, A., Schreuder, R., and Frauenfelder, U.H. (1989). Grapheme context effects on phonemic processing. Lang. Speech *32*, 89–108.

Dijkstra, T., Frauenfelder, U.H., and Schreuder, R. (1993). Bidirectional grapheme-phoneme activation in a bimodal detection task. J. Exp. Psychol. Hum. Percept. Perform. *19*, 931–950.

Duvernoy, H.M. (1999). The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI and Blood Supply, Second Edition (Wien, NY: Springer-Verlag).

Fischl, B., Sereno, M.I., Tootel, R.B.H., and Dale, A.M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. *8*, 272–284.

Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., and Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron *40*, 859–869.

Fort, A., Delpuech, C., Pernier, J., and Giard, M.H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb. Cortex *12*, 1031–1039.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In Surface Dyslexia, K.E. Patterson, J.C. Marshall, and M. Coltheart, eds. (London: Routledge & Kegan-Paul), pp. 301–330.

Frost, R., Repp, B.H., and Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? J. Mem. Lang. *27*, 741–755.

Gauthier, I., Tarr, J., Moylan, J., Skudlarski, P., Gore, C., and Anderson, W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. J. Cogn. Neurosci. *12*, 495–504.

Genovese, C., Lazar, N., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage *15*, 870–878.

Gros, H., Boulanouar, K., Viallard, G., Cassol, E., and Celsis, P. (2001). Event-related functional magnetic resonance imaging study of the extrastriate cortex response to a categorically ambiguous stimulus primed by letters and familiar geometric figures. J. Cereb. Blood Flow Metab. *21*, 1330–1341.

Hadjikhani, N., and 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.

Hardy, M.H., Smythe, P.C., Stennet, R.G., and Wilson, H.R. (1972). Developmental patterns in elemental reading skills: phoneme-grapheme and grapheme-phoneme correspondences. J. Educ. Psychol. *63*, 433–436.

Hikosaka, K., Iwai, E., Saito, H., and Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J. Neurophysiol. *60*, 1615–1637.

Hinton, G.E., McClelland, J.L., and Rumelhart, D.E. (1986). Distributed representations. In Parallel Distributed Processing: Explorations in the Microstructutre Of Cognition, D.E. Rumelhart and J.L.

McClelland, eds. (Cambridge, Massachusetts: The MIT Press), pp. 77–109.

Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H.J. (2002). Phonetic perception and the temporal cortex. Neuroimage *15*, 733–746.

Jones, J.A., and Callan, D.E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. Neuroreport *14*, 1129–1133.

Kaas, J.H., and Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. USA *97*, 11793–11799.

Kim, J.-J., Crespo-Facorro, B., Andreasen, N.C., O'Leary, D.S., Zhang, B., Harris, G., and Magnotta, V.A. (2000). An MRI-based parcellation method for the temporal lobe. Neuroimage *11*, 271–288.

King, A.J., and Calvert, G.A. (2001). Multisensory integration: perceptual grouping by eye and ear. Curr. Biol. *11*, R322–R325.

Liberman, A.M. (1992). The relation of speech to reading and writing. In Orthography, Phonology, Morphology and Meaning, R. Frost and L. Katz, eds. (Amsterdam: Elsevier Science Publishers B.V.), pp. 167–178.

Longcamp, M., Anton, J.L., Roth, M., and Velay, J.L. (2003). Visual presentation of single letters activates a premotor area involved in writing. Neuroimage *19*, 1492–1500.

Macaluso, E., Frith, C.D., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. Science *289*, 1206–1208.

Massaro, D.W., Cohen, M.M., and Thompson, L.A. (1988). Visible language in speech perception: lipreading and reading. visible language *22*.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature *263*, 747.

Meredith, M.A. (2002). On the neuronal basis for multisensory convergence: a brief overview. Cogn. Brain Res. *14*, 31–40.

National Reading Council. (1998). Preventing Reading Difficulties in Young Children (Washington, DC: National Academy Press).

Olson, I.R., Christopher Gatenby, J., and Gore, J.C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Cogn. Brain Res. *14*, 129–138.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N.A., De Giovanni, U., Sensolo, S., and Fazio, F. (2003). A functional-anatomical model for lipreading. J. Neurophysiol. *90*, 2005–2013.

Perfetti, C.A., and Sandak, R. (2000). Reading optimally builds on spoken language: Implications for deaf readers. J. Deaf Stud. Deaf Educ. *5*, 32–50.

Polk, T., Stallcup, M., Aguirre, G., Alsop, D., D'Esposito, M., Detre, J., and Farrah, M. (2002). Neural specialization for letter recognition. J. Cogn. Neurosci. *14*, 145–159.

Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. Neuron *28*, 617–625.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O.V., Lu, S.T., and Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci. Lett. *127*, 141–145.

Schroeder, C.E., and Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn. Brain Res. *14*, 187–198.

Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. Neurosci. Res. *47*, 277–287.

Shah, N., Steinhoff, S., Mirzazade, S., Zafiris, O., Grosse-Ruyken, M.L., Jäncke, L., and Zilles, K. (2000). The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. Neuroimage *12*, 100–108.

Sumby, W.H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. *26*, 212–215.

Suzuki, M., Kitano, H., Kitanishi, T., Itou, R., Shiino, A., Nishida, Y., Yazawa, Y., Ogawa, F., and Kitajima, K. (2002). Cortical and subcortical activation with monaural monosyllabic stimulation by functional MRI. Hear. Res. *163*, 37–45.

Tervaniemi, M., Medvedev, S.V., Alho, K., Pakhomov, S.V., Roudas, M.S., van Zuijen, T.L., and Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. Hum. Brain Mapp. *10*, 74–79.

Wallace, M.T., Wilkinson, L.K., and Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. J. Neurophysiol. *76*, 1246–1266.

Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb. Cortex *13*, 1034–1043.