Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex

EJN
European Journal of Neuroscience

European Journal of Neuroscience, Vol. 28, pp. 500–509, 2008 doi:10.1111/j.1460-9568.2008.06350.x

Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex

Vera Blau,[1] Nienke van Atteveldt,[1] Elia Formisano,[1] Rainer Goebel[1] and Leo Blomert[2]
[1]Department of Cognitive Neuroscience, Faculty of Psychology & Neuroscience (FPN), University of Maastricht, 6200 MD Maastricht, The Netherlands
[2]Faculty of Psychology, Maastricht Brain Imaging Center (M-BIC), Maastricht, The Netherlands

Keywords: fMRI, integration, multisensory, noise

Abstract

Letters and speech sounds are the basic units of correspondence between spoken and written language. Associating auditory information of speech sounds with visual information of letters is critical for learning to read; however, the neural mechanisms underlying this association remain poorly understood. The present functional magnetic resonance imaging study investigates the automaticity and behavioral relevance of integrating letters and speech sounds. Within a unimodal auditory identification task, speech sounds were presented in isolation (unimodally) or bimodally in congruent and incongruent combinations with visual letters. Furthermore, the quality of the visual letters was manipulated parametrically. Our analyses revealed that the presentation of congruent visual letters led to a behavioral improvement in identifying speech sounds, which was paralleled by a similar modulation of cortical responses in the left superior temporal sulcus. Under low visual noise, cortical responses in superior temporal and occipito-temporal cortex were further modulated by the congruency between auditory and visual stimuli. These cross-modal modulations of performance and cortical responses during an unimodal auditory task (speech identification) indicate the existence of a strong and automatic functional coupling between processing of letters (orthography) and speech (phonology) in the literate adult brain.

Introduction

In alphabetic scripts, learning the correspondences between the smallest units of written and spoken language (letters/graphemes vs. speech sounds/phonemes) is a key step in reading acquisition (Ehri, 2005). A failure to acquire sufficiently automatized letter–speech sound associations has been proposed as a cause for dyslexia (Vellutino *et al.*, 2004). Whereas behavioral studies suggest the fast and automatic cross-modal activation from letter to sound representations (Dijkstra *et al.*, 1989, 1993; Frost & Katz, 1989; Borowsky *et al.*, 1999), it is not clear which neural mechanisms subserve the influences from letters to speech sounds and its effects on behavior in literate adults.

Several neuroimaging studies have investigated the neural basis of letter-to-sound conversion both at the level of word processing (Fiez *et al.*, 1999; Tagamets *et al.*, 2000; Booth *et al.*, 2002; Fiebach *et al.*, 2002; Simos *et al.*, 2002; Booth, 2007) and at the level of letters and speech sounds (Raij *et al.*, 2000; van Atteveldt *et al.*, 2004; Hashimoto & Sakai, 2004). Converging results indicate the involvement of heteromodal superior temporal cortex (STS/STG), auditory Heschl sulcus/planum temporale (HS/PT) and visual occipito-temporal cortex (OT). The processing of congruent vs. incongruent letters–speech sound pairs was found to enhance cortical activity in superior temporal and auditory cortex (van Atteveldt *et al.*, 2004), indicating the modulation of neural responses specific for learned audiovisual associations.

Correspondence: V. Blau, as above.
E-mail: v.blau@psychology.unimaas.nl

Received 22 February 2008, revised 19 May 2008, accepted 3 June 2008

The aim of the present behavioral and functional magnetic resonance imaging (fMRI) study was twofold. First, we investigated whether letters influence speech processing even if visual processing is not explicitly required by the task. Second, we investigated whether the neural influences of congruent/incongruent visual stimulation on speech sound processing are relevant for the behavioral facilitation/inhibition of speech identification.

The first question was motivated by behavioral and electrophysiological findings indicating that, in literate adults, the access from letter to sound representations is fast and highly automated (Dijkstra *et al.*, 1989, 1993; Proverbio *et al.*, 2004; Froyen *et al.*, 2008; Perre & Ziegler, 2008). Hence, we presented visual letters as task-irrelevant stimuli during a unimodal forced choice speech identification task. Note that analogous to earlier reports (Dijkstra *et al.*, 1989) the term 'automatic' is used to indicate cross-modal influences that occur even though the visual letter is neither directly task-relevant nor connected to the required response. Similar designs have been used in electrophysiology research to investigate the influence of task-irrelevant auditory stimulation on visual processing (McDonald *et al.*, 2003; Busse *et al.*, 2005). Speech sounds were presented alone or bimodally in congruent and incongruent combinations with the letter. It was predicted that letters modulate the processing of speech sounds in heteromodal (STS/STG) and/or 'modality-specific' cortices (HS/PT, OT), indicating automatic visual influences on speech identification.

The second question was motivated by the observation that previous neuroimaging studies have often either used passive perception tasks (Calvert *et al.*, 2000; Olson *et al.*, 2002; van Atteveldt *et al.*, 2004, 2007) or indirect performance measures (Callan *et al.*, 2003; Macaluso *et al.*, 2004), concealing the contribution of audiovisual integration to

behavioral response facilitation/inhibition. We addressed this issue by presenting visual letters using three degradation levels (low, medium and high visual noise), and examining the consequences of this manipulation at the behavioral and neural level. We predicted parametric influences of visual degradation not only for behavioral responses to speech sounds but also for neural activity related to letter–speech sound processing.

Materials and methods

Subjects

Nineteen healthy, right-handed subjects (eight male) with normal hearing and normal or corrected-to-normal vision were recruited for the present fMRI experiment. All participants were native Dutch speakers enrolled in (under)graduate programs at Maastricht University (mean age: 21.4 years; SD: 3.5). The psychophysical experiment included 10 right-handed subjects with normal or corrected-to-normal vision (three male) drawn from the same population of students (mean age: 20.5 years; SD: 1.65). All subjects were screened to exclude psychiatric, neurological or reading disorders. Prior to testing subjects gave informed written consent to participate. Subjects were paid for their participation, which was voluntary and in accordance with the Faculty of Psychology ethical guidelines.

Stimuli

Stimuli were visual letters and auditory speech sounds corresponding to the Dutch vowels [a] and [e]. Auditory stimuli were created by digitally recording the voice of a native Dutch female speaker at a sampling rate of 44.1 kHz (16 bit). All recordings were resampled at 22.05 kHz offline and had a total length of approximately 400 ms. Speech sounds were then mixed with pink auditory noise (sampling frequency 22 kHz, 16 bit) of the same length using GOLDWAVE v5.12 (GoldWave). Subsequently, combined auditory stimuli were matched for overall intensity (±74 dB). Behavioral pilot-testing indicated how subjects perceived the degraded speech sounds using various noise levels (5, 10, 15, 20 and 25 dB), one of which was selected for the present fMRI study (10 dB).

Visual stimuli were lower-case letters (print-type: times new roman; visual angle: 2.8 × 3.8 degrees) corresponding to vowels 'a' and 'e'. All visual letters were in a white font on a gray background. Using PaintShopPro (Corel Cooperation), noise in the visual domain was added by applying a two-step procedure: first, the overall contrast was lowered by 40% in order to remove high-contrast components that would prohibit sufficient visual masking; second, a visual mask was overlaid on top of the letter using three opacity levels. The mask was created by taking bar-shaped figures from the original low-contrast stimulus and randomly reassembling those shapes to create an abstract visual profile. This way visual letters were created with low, medium and high levels of noise corresponding to good, medium and low letter perceptibility (Fig. 1A).

Design and task

Stimuli were presented in unimodal auditory and audiovisual conditions on a trial-by-trial basis within a forced-choice auditory identification task. During cross-modal trials auditory and visual stimuli were presented simultaneously in either congruent or incongruent combinations (congruency factor). Auditory stimuli were presented slightly degraded (see 'Stimuli' section) to prevent behavioral ceiling effects. Additionally, the visual letter had high, medium or low quality (noise factor) resulting in a total of seven experimental conditions: unimodal auditory (A); audiovisual congruent with low amounts of visual noise (AVcon1); medium amounts of visual noise (AVcon2); and high amounts of visual noise (AVcon3). Likewise, audiovisual incongruent trials were presented with low amounts of visual noise (AVinc1); medium amounts of visual noise (AVinc2); and high amounts of visual noise (AVinc3).

A psychophysical study, using an unrelated subject group, preceded the fMRI study (in Results referred to as 'outside scanner') and consisted of two runs with 40 trials per condition resulting in a total of 280 trials in pseudorandom order, which were repeated with a delay of 400 ms. Subjects were instructed to identify via button-press whether they perceived the speech sound /a/ or /e/. In addition to the overt auditory task, subjects were instructed to maintain fixation on the center of the screen. To maximize the assumed cross-modal perceptual gain induced by visual co-stimulation during speech identification, we used a mixed experimental design, in which each experimental run contained congruent and incongruent trials presented at a ratio of 75% congruent : 25% incongruent trials in one half of the run and 75% incongruent : 25% congruent trials in the other half of the same run (no blocked presentation). The order of 75% congruent vs. 75%

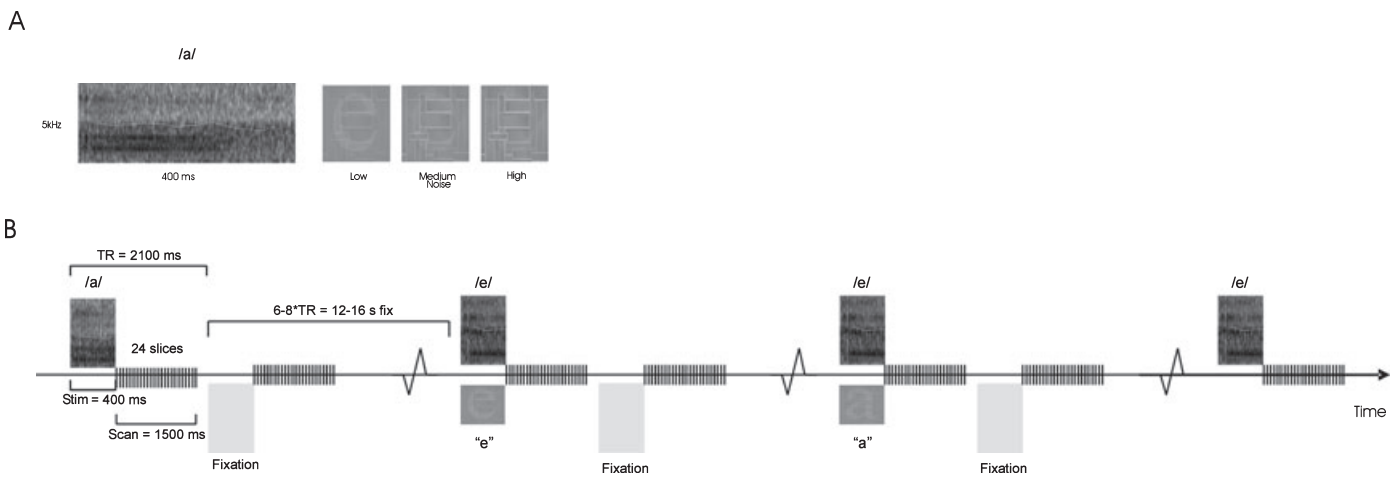

FIG. 1. (A) Exemplars of experimental stimuli, auditory /a/ and visual 'e' masked with low, medium and high degrees of visual noise. (B) Illustration of MR experimental setup with timing parameters of four representative trials.

incongruent periods within a run was counterbalanced across subjects. No external cue indicated a switch in congruent : incongruent ratio in line with subjects' consistent report of being unaware of the clustered trial structure. The necessity of this manipulation was established in a pilot experiment. Its rationale can be described as follows: the identity of the speech sounds should not be predictable by attending the visual letter only, as in the case of blocked stimulus presentation. At the same time a 50% congruent : incongruent ratio would trigger a situation in which it is unclear whether subjects will use the visual channel in order to solve the speech identification task or whether they will try to reduce its influence by ignoring it. Hence, a design investigating the perceptual gain on speech processing experienced through the presence of visual letters has to provide a clear incentive for processing visual information. This was achieved by changing the congruent : incongruent ratio to 75% : 25% in the present study, a situation where subjects profit from processing the visual letter for speech sound identification without making the decision predictable.

To validate the current paradigm, behavioral responses between congruent trials embedded in 75% congruent vs. 75% incongruent trials were compared (and incongruent trials, respectively). No difference in accuracy of speech identification was induced as the trial ratio changed, indicating the absence of strategic task differences or global attention/arousal effects and supporting the feasibility of the current paradigm for investigating visual influences on speech identification.

The fMRI experiment, during which behavioral data were also acquired (in Results referred to as 'inside scanner') differed from the psychophysical experiment in the following way: one session consisted of two runs with 24 trials per condition and each experimental run consisted of widely spaced single trials (6–8 TRs or 12–16 s). Aside from this adjustment in the temporal spacing of events, auditory speech sounds were presented slightly less degraded during scanning (8 dB) to account for the auditory saturation effects created by the scanner environment.

Image acquisition and data analysis

Data were acquired on a 3T Siemens Allegra head scanner using gradient echo planar imaging (EPI). All participants completed two experimental runs, in which 582 T2*-weighted BOLD contrast volumes were acquired for each run. Based on structural information from a previous 9-slice localizer scan, 24 axial slices were positioned in each subject to cover the auditory and visual cortices (TE = 30 ms; TRslice = 63 s; matrix size: $64 \times 64 \times 24$; FOV: 192 mm^2; voxel size: $3 \times 3 \times 4$ mm^3). Volume acquisition time was 1.5 s followed by a silent delay of 600 ms in which stimuli were presented resulting in a TR of 2.1 s (Fig. 1A). The inter-scan gap was used to minimize the effects of scanning noise on experimental activation (Jäncke *et al.*, 2002). Furthermore, the use of a long intertrial interval (mean: 7 TR/14.7 s; jitter: ±1 TR/2.1 s) permitted to separate the hemodynamic response to individual trials with the advantage of modeling BOLD signal changes on a single trial basis. A high-resolution T1-weighted anatomical image (voxel size: $1 \times 1 \times 1$ mm^3) was acquired for each subject using a three-dimensional magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (TR = 2.3 s; TE = 3.39 ms; matrix size: 256×256; 192 slices/slab).

Imaging data were analysed using BrainVoyager QX software (Brain Innovation, Maastricht, Netherlands; Goebel *et al.*, 2006). Functional data were preprocessed to correct for slice scan time differences (sinc interpolation), three-dimensional motion artifacts (trilinear interpolation), linear drifts and low-frequency non-linear drifts due to head motion (high pass filter ≤ 3 cycles/time course). No spatial or additional temporal smoothing was applied. Functional data were then coregistered with the anatomical volume and transferred into standard stereotaxic space using Talairach normalization (Talairach & Tournoux, 1988). A design matrix was set up to model all experimental conditions for each acquisition. Error trials were modeled as separate predictor and excluded from further analysis. Statistical maps were generated by modeling the evoked hemodynamic response for the different stimuli as boxcars convolved with a two-gamma hemodynamic response function in the context of the general linear model (GLM). Population-level inferences concerning BOLD signal changes between the experimental conditions were based on a random effects model with predictors separated for each subject. Two different GLMs were used for the present analysis.

The first GLM (GLM1) was a single-factor model including all seven conditions as separate predictors. This model was used to localize brain areas that responded to both unimodal auditory and audiovisual conditions (referred to as 'conjunction analysis', see Fig. 3). Overlapping activation for unimodal and cross-modal conditions is a prerequisite for investigating the additional value of adding a visual stimulus. Therefore, only voxels that were significantly activated in the conjunction analysis were included in further analysis by application of a functional mask. The conjunction was computed in the following way: [Auditory > Baseline] n [AudioVisual > Baseline]. In this analysis, a new statistical value is computed for each voxel as the minimum of the statistical values resulting from the two included contrasts. Activations were considered significant after correcting for multiple comparisons using cluster-size thresholding (Forman *et al.*, 1995; Goebel *et al.*, 2006). This method exploits the fundamental assumption that areas of activity tend to stimulate signal changes over spatially contiguous groups of voxels rather than over sparsely isolated voxels. The computation of the minimum cluster threshold is accomplished via MonteCarlo simulation of the random process of image generation, followed by the injection of spatial correlations between neighboring voxels, voxel intensity thresholding, masking (optional) and cluster identification. Starting from a manually adjusted voxel-level probability threshold, a minimum cluster size threshold is automatically set yielding 5% (or less) protection against false positive detection at the cluster level. In the present study, an initial voxel-level (uncorrected) threshold was set to 0.001 ($t = 4.25$) uncorrected, resulting in a cluster-level of 160 mm^3 after 1000 iterations, corresponding to a corrected false positive probability of 5% or less. Furthermore, GLM1 was used in a *post hoc* analysis to investigate the effect of congruency for all three levels of visual noise separately by using the contrasts [AV1con > AV1inc], [AV2con > AV2inc] and [AV3con > AV3inc] (referred to as 'congruency analysis'; Fig. 5). The correction for multiple comparisons was again carried out using cluster size thresholding. To investigate the congruency effect, we initially set an uncorrected *P*-value of 0.01 ($t = 3.27$), which resulted in a cluster size threshold of 149 mm^3 after 1000 iterations (alpha < 0.05).

The second GLM (GLM2) was a 2×3 factorial model with factors congruency (congruent, incongruent) and visual noise level (low, medium, high) in order to specifically search for interaction effects (referred to as 'interaction analysis'; Fig. 4). We corrected for cluster size using an initial *P*-value of 0.01 ($t = 3.79$), which after 1000 iterations resulted in a cluster size threshold of 111 mm^3 (alpha < 0.05).

Statistical analyses for selected regions of interest in GLM1 and GLM2 were based on average percent signal change values within a region of interest (ROI). Peak percent signal change values were also

used but were plotted for visualization purposes only (bar graphs: Figs 4 and 5).

Results

Behavioral results outside scanner (psychophysics)

Behavioral outcomes from the psychophysical experiment outside the scanner revealed that the identification of speech sounds was influenced by the presentation of a task-irrelevant visual letter. To assess the influence of the visual letter on speech sound identification, unimodal auditory responses were subtracted from responses in audiovisual conditions for each participant (AV-A). Positive reaction time difference scores therefore indicated disrupted speech sound identification on audiovisual trials vs. auditory trials, whereas negative difference scores indicated enhanced identification performance. Based on the difference scores for each cross-modal condition, we found a significant congruency by visual noise interaction on reaction times for the identification of the speech sounds ($F_{1.42,9} = 14.91$, $P < 0.001$; Fig. 2, upper panel). Comparisons of congruent vs. incongruent letter–speech sound pairs revealed the strongest influence of audiovisual congruency under the condition where the letter was least degraded (AV1) ($t_9 = –5.01$, $P < 0.001$) and decreasing, but still significant congruency effects for medium (AV2) ($t_9 = –2.64$, $P < 0.05$) and high amounts of visual noise (AV3) ($t_9 = –2.55$, $P < 0.05$). Thus, subjects' identification speed of speech sounds was maximal when accompanied by a visual letter that was minimally distorted and matched the speech sound. In contrast, identification was slowest when accompanied by a minimally distorted visual letter that did not match the sound. This means that speech sound identification on audiovisual incongruent trials relative to auditory alone trials was disrupted, whereas speech sound identification on audiovisual congruent trials relative to auditory alone trials was enhanced. Both reaction time effects, enhancement and disruption, were weakened when the visual letter became more degraded. Furthermore, subjects in the behavioral study also identified speech sounds more accurately when these sounds were combined with well-perceivable letters in congruent combinations, relative to their accuracy for incongruent letter–sound combinations ($t_9 = 3.53$, $P < 0.01$). This accuracy effect was smaller but remained significant at the medium level of noise ($t_9 = 2.69$, $P < 0.05$), and was absent at high amounts of noise. This pattern was reflected in a significant interaction between visual noise and letter-sound congruency ($F_{1.42,9} = 5.88$, $P < 0.05$).

Behavioral results inside scanner

Similar to the behavioral results outside the scanner, we found a significant noise by congruency interaction effect ($F_{2,18} = 15.6$, $P < 0.001$) during scanning. Participants responded faster to speech sounds when the sounds were accompanied by a low-noise, congruent visual letter relative to an incongruent letter ($t_{18} = –4.67$, $P < 0.00$). This response enhancement was also present at medium visual noise levels ($t_{18} = –3.47$, $P < 0.01$), but absent at high visual noise levels. In other words, at low and medium visual noise levels, congruent audiovisual trails enhanced the identification of speech sounds relative to auditory alone trials, whereas audiovisual incongruent trials disrupted speech sound identification relative to auditory alone trials. This result indicates that subjects process the visual letters effectively also during incongruent trials and, hence, discard an explanation of the congruency effect in terms of general differential attention. Moreover, letter–sound congruency and visual noise level also interacted to predict the accuracy for speech sound identification ($F_{1.22,18} = 9.26$, $P < 0.01$; Fig. 2, lower panel). Pair-wise comparisons assessing the

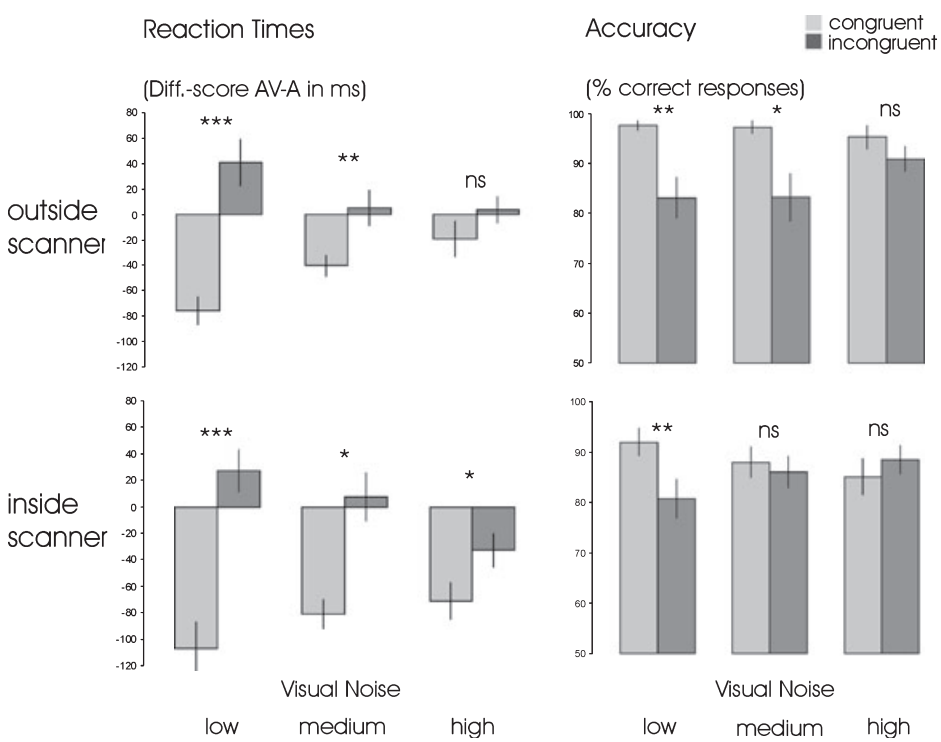

FIG. 2. Bar diagrams of reaction time difference scores ([AV-A] left panel) and accuracy data (right panel) for behavioral identification of speech sounds for the two independent subject populations outside (upper panel) and inside the scanner (lower panel) reveal remarkably similar patterns of performance. Stars indicate the level of significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

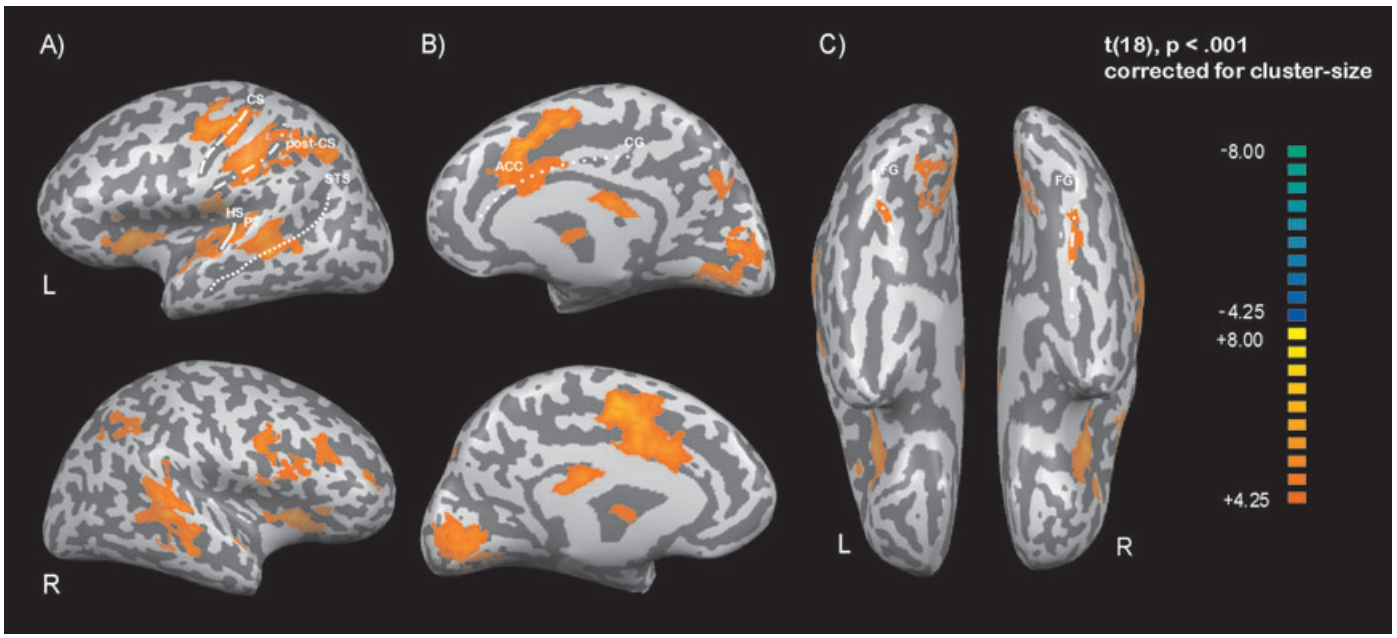

FIG. 3. Group results for the conjunction analysis (cluster size corrected at alpha < 0.05) projected on inflated cortical surface of one representative subject. Lateral (A), medial (B) and ventral (C) projections depict overlapping regions between auditory and audiovisual conditions. CG, cingulate gyrus; CS, central sulcus; FG, fusiform gyrus; HS, Heschl sulcus; PT, planum temporale; STS, superior temporal sulcus.

effect of congruency per level of visual noise showed that the effect was mainly driven by more accurate responses during congruent vs. incongruent trials when the visual letter contained low noise (AVcon1 > AVinc1) ($t_{18} = 3.38$, $P < 0.01$). In contrast, when the visual letter contained medium or high levels of noise, congruency between letters and speech sounds did not significantly contribute to more accurate sound identification. Overall, the pattern of reaction time and accuracy effects outside the scanner were quantitatively and qualitatively similar to behavioral performance inside the scanner (Fig. 2).

MRI results

Conjunction analysis

Figure 3 provides an overview of brain areas activated by unimodal and cross-modal stimulation, as revealed by the conjunction analysis [Auditory > Baseline] n [AudioVisual > Baseline]. Only those voxels that were significantly activated in this analysis were included in further analysis. Our goal was to first compare the global pattern of activation to previous passive (van Atteveldt *et al.*, 2004, 2007) studies on letter–sound integration. Secondly, the degree of overlapping activation for unimodal auditory stimulation and audiovisual stimulation is a first indication for potential 'additional' activation changes induced by the visual stimulus. We found areas of activation along the middle portions of the STG/STS extending into the middle temporal gyrus (MTG) bilaterally. Furthermore, activated brain regions included HS/PT in the left hemisphere and visual striate and extrastriate brain regions bilaterally, pre- and postcentral gyri and inferior parietal cortex bilaterally as well as basal ganglia, insular, putamen and left anterior cingulate cortex (Fig. 3).

Interaction analysis

The second step in the analysis was to directly assess how the visual noise level manipulation influenced speech sound processing, and to assess potential interactions between congruency and noise level via

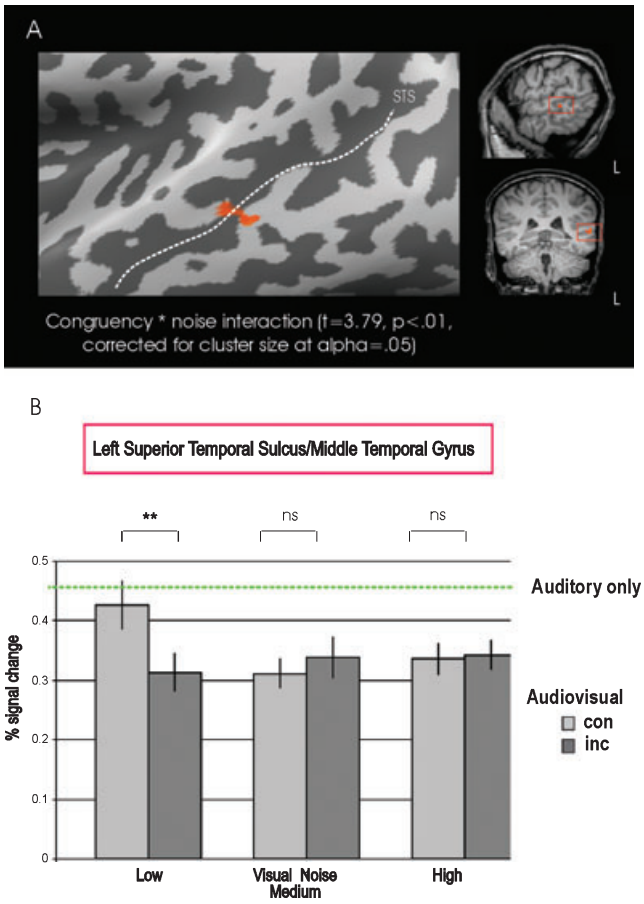

FIG. 4. Group results for congruency × noise interaction analysis (GLM2) corrected for cluster size (alpha < 0.05), projected on anatomical scan (left) and inflated cortical surface (right) of one representative subject (A) and corresponding peak BOLD percent signal change values in left STS/MTG region of interest (B) (**$P < 0.01$, also see Table 1).

a 2 × 3-factorial GLM (GLM2, see Materials and methods). No main effects of visual noise level or congruency were observed. A significant interaction of letter–sound congruency and visual noise was found in the left STS extending into the MTG (Fig. 4A; $P < 0.05$). *Post hoc* ROI GLM analysis indicated that this interaction reflected a congruency effect in the low visual noise condition ($t_{18} = 2.6$, $P < 0.05$) and that this effect was not significant for the medium- and high-noise levels (Fig. 5B). To further explore the nature of the significant congruency effect in the low-noise condition, direct comparisons of STS/MTG responses to congruent and incongruent letter–sound pairs at low visual noise vs. the unimodal auditory condition were performed. A significant reduction in activity for the incongruent vs. the auditory condition was revealed ($t_{18} = –2.56$, $P < 0.05$). Interestingly, the pattern of fMRI responses shown in the congruency by noise interaction resembled the behavioral interaction effect in that stronger STS/MTG signal changes corresponded to higher accuracy and speedier reaction times during low-noise audiovisual trials in the congruent condition (see behavior results). However, whereas the finding of an interaction effect on speech identification was due to parametric influences of visual noise on letter–speech sound congruency, cortical responses in STS/MTG exhibited rather stepwise instead of gradual influences of visual noise. This lack of a parametric effect on STS/MTG responses may be due to many factors, including the weakened behavioral effects during scanning and the indirect nature of the BOLD signal.

Post hoc congruency contrasts

We further contrasted audiovisual congruent and incongruent conditions for all noise levels separately in order to explore the congruency effects per noise level on a whole brain level. This analysis is free from the constraint that activity has to vary as a function of visual noise and can therefore be considered more liberal. We found congruency effects within the low visual noise condition (AV1con vs. Av1inc) only, as shown in Fig. 5. Activated brain areas included a region of the anterior left STG, STS including medial MTG, and portions of the fusiform gyrus (FG) bilaterally in proximity to what has been termed the visual word form area (Cohen *et al.*, 2002). A region in HS/PT (Fig. 4) was additionally activated. The left HS/PT ROI did not survive the cluster size criterion of 149 activated voxels; however, in agreement with previous findings, this activation is reported at a smaller cluster size (82 voxels). Finally, stronger activation for congruent vs. incongruent letter–sound combinations was found in the precentral gyri bilaterally and in the left putamen. ROI-based GLMs further indicated that activation to congruent letter–sound pairs was stronger in all selected ROIs than activation related to the presentation of speech sounds in isolation [left anterior STG: $t_{18} = 2.9$, $P < 0.01$; left MTG/STS: $t_{18} = 3.3$, $P < 0.01$; FG: $t_{18} = 2.7$, $P < 0.05$ (left); $t_{18} = 3.9$, $P < 0.01$; HS/PT: $t_{18} = 4.2$, $P < 0.01$ (left); $t_{18} = 2.7$, $P < 0.05$ (right)]. In contrast, responses to incongruent combinations of letters and speech sounds did differ from speech sounds in isolation only in left STS/MTG and right HS/PT; in these areas the BOLD response to incongruent letter–speech sound pairs was reduced compared with speech sounds alone ($t_{18} = 5.3$, $P < 0.001$; HS/PT: $t = 2.55$, $P < 0.05$). A complete listing of the results reported in the different analyses is shown in Table 1.

Discussion

The first goal of the present study was to investigate the influence of task-irrelevant visual letters on the processing of speech sounds. In order to address this question we chose an experimental design in which visual letters were presented as task-irrelevant stimuli during unimodal speech sound identification. We found visual influences on speech processing in several brain regions, including the superior temporal cortex, auditory cortex and extrastriate visual cortex. Secondly, we investigated the behavioral relevance of the neural integration of letters and speech sounds for speech identification. In order to address this question, we manipulated the quality of the visual letter parametrically. Our results demonstrate similar effects of visual stimulus quality on the neural processing of letter–speech sound pairs in STS/MTG and on the behavioral identification of speech.

Brain regions supporting the visual influence on speech sound processing

Cortical responses in a network of brain regions including STS, as well as 'modality-specific' auditory (HS/PT; aSTG) and visual (FG) association cortex were influenced by the co-presentation of visual letters during speech identification. In the absence of any explicit requirement to process the letters, those regions responded more vigorously to congruent as opposed to incongruent bimodal stimulation. Analogously, as indicated by accuracy and reaction time data, speech sound identification was facilitated in the context of congruent letters and inhibited in the context of incongruent letters. In line with these data, previous behavioral studies found (Dijkstra *et al.*, 1989, 1993) that the presentation of congruent syllable primes facilitated the detection of a cross-modal target.

STS is a well-known candidate structure for the integration of audiovisual (and tactile) information as it receives input from multiple senses via cortical and subcortical connections (Beauchamp *et al.*, 2004; Macaluso *et al.*, 2004). Our finding of a modulation of STS responses to speech sounds by letters indicates cross-modal visual influences in this region, consistent with previous reports on passive letters–sound integration (van Atteveldt *et al.*, 2004, 2007) and the integration of cross-modal speech during active and passive tasks (Calvert *et al.*, 2000; Raij *et al.*, 2000; Olson *et al.*, 2002; Wright *et al.*, 2003). Moreover, the location of the present STS result closely matched that of previous neuroimaging investigations on letters and speech sounds (van Atteveldt *et al.*, 2004, 2007) and audiovisual speech (Sekiyama *et al.*, 2003; Beauchamp *et al.*, 2004).

Interestingly, the congruency effect in STS was mainly characterized by a stronger suppression of incongruent than congruent letters–speech sound pairs compared with speech sounds in isolation. One potential explanation for this finding might be that the fairly strong responses to unimodal auditory stimulation in the present tasks increased the likelihood for finding multisensory suppression at the cost of finding multisensory enhancement effects. Monkey electrophysiological work has shown that a strong unimodal responsiveness of neurons allows for relatively less enhancement of cortical firing than a weak unimodal response (Stanford *et al.*, 2005). In the present study, this auditory saturation might be explained by the unimodal auditory task with corresponding allocation of attention to the auditory modality.

More importantly, however, the effect of congruency on STS responses was elicited by task-irrelevant visual letters, indicating the automatic cross-modal modulation of STS responses to speech sounds. This finding indicates the mapping between orthography and phonology in literate adults.

Cross-modal visual influences on speech processing were furthermore observed in regions of the posterior (HS/PT; see Fig. 5) and anterior (aSTG; see Fig. 5) auditory association cortex. HS/PT

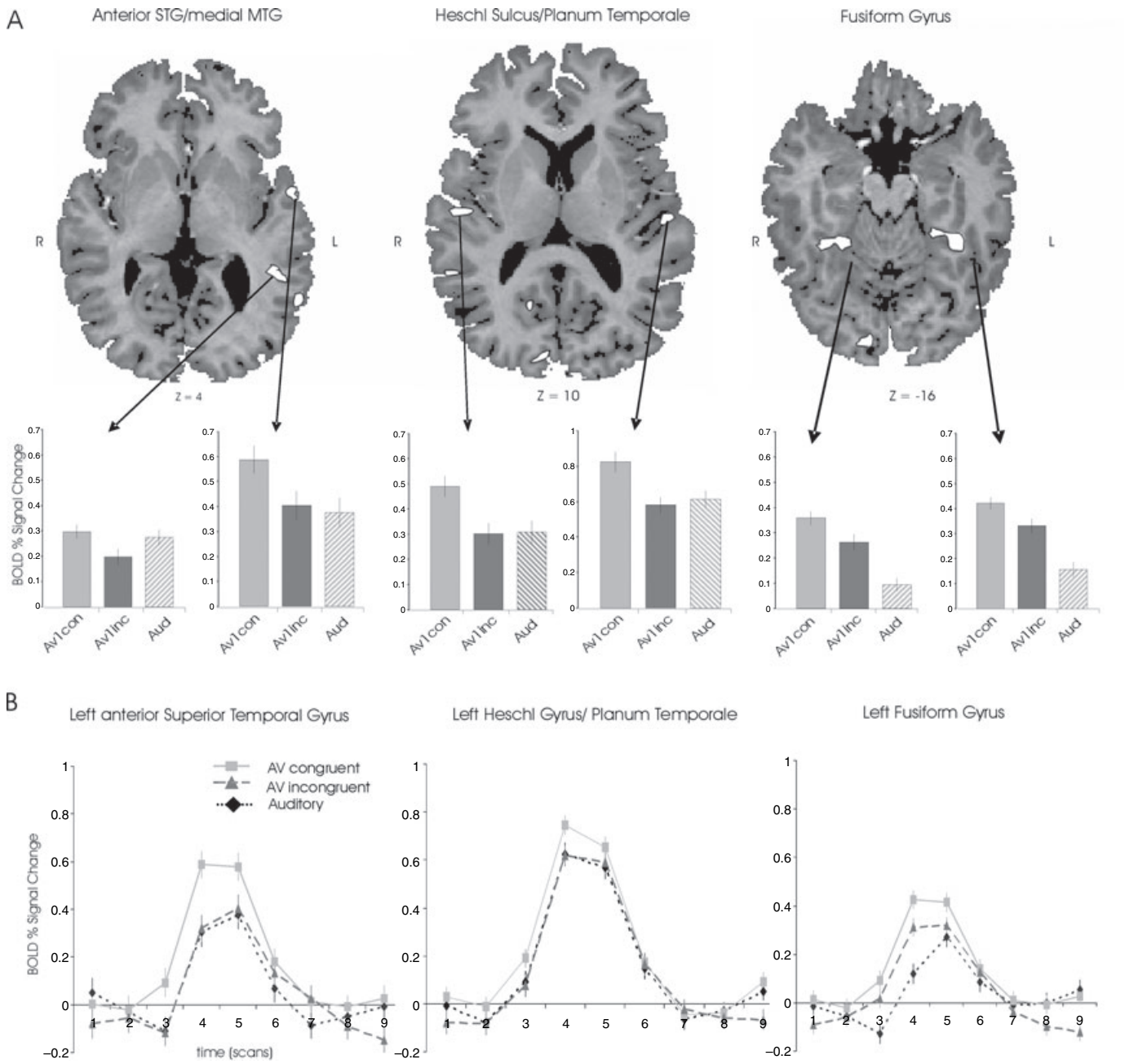

FIG. 5. Group results for the congruency contrast (GLM1: AV1con > AV1inc). (A) Loci of activation are shown for three regions of interest: left anterior superior temporal gyrus (STG) and left STS/middle temporal gyrus (MTG) (left), left and right HS/PT (middle), and FG bilaterally (right). The corresponding peak percent signal change values are shown for audiovisual congruent (light gray bars), incongruent (dark gray bars) and auditory (stripes) conditions (Table 1). (B) Example BOLD response time courses for three regions of interest in the left hemisphere (also see Table 1).

activation is generally associated with auditory, pre-phonemic analysis of complex sounds (Binder *et al.*, 1996; Seifritz *et al.*, 2002), but has also been found for the integration of spoken and written language (Nakada *et al.*, 2001; van Atteveldt *et al.*, 2004) and the acquisition of new audiovisual associations (Hasegawa *et al.*, 2004; Hashimoto & Sakai, 2004). Before learning to read, letters and speech sounds represent new, artificial audiovisual associations that have to be explicitly learned. PT might provide the necessary link between sensory representations of letters and speech sounds with motor representations involved in reading/speaking (Wilson *et al.*, 2004) and writing (Longcamp *et al.*, 2003). Anterior temporal lobe responses, in comparison, are more closely associated with the processing of intelligible than non-intelligible speech (Scott *et al.*, 2000; Scott & Johnsrude, 2003; Hwang *et al.*, 2006), and have even been explicitly linked to the processing and categorization of vowels (Obleser *et al.*, 2006). Given that information about the identity of the speech sound is presumably already available in aSTG, congruency effects of visual letters on the processing of speech sounds might be expected in this region. This interpretation is in accordance with neurofunctional models of speech perception, which have implicated anterior parts of STS and STG as being part of a 'what' pathway for cross-modal object recognition (Sestieri *et al.*,

TABLE 1. ROI details and statistics per analysis

Brain area	Hemisphere	Talairach coordinates			Voxels (*n*)	Effect size*		Pairwise comparisons		
		x	*y*	*z*		*t*-value	*P*-value	Con > A	Inc > A	Con > Inc
Interaction test used for ROI selection†										
STS	Left	−58	−37	7	98	3.23	0.0067	0.93	0.04	0.03
Congruency test used for ROI selection†										
FG	Left	−35	−42	−16	286	3.20	0.0008	0.03	0.66	0.00
FG	Right	32	−43	−17	254	3.08	0.0066	0.00	0.64	0.00
aSTG	Left	−56	−1	4	149	3.31	0.0041	0.01	0.06	0.00
STS/MTG	Left	−47	−44	2	169	3.25	0.0048	0.01	0.00	0.00
HS/PT	Left	−54	−19	12	82	5.51	0.0012	0.00	0.10	0.00
HS/PT	Right	−50	−15	13	147	4.99	0.0011	0.02	0.03	0.00
PCG	Left	−47	−27	39	285	2.86	0.0123			
PCG	Right	46	−28	41	219	3.05	0.0009			
Putamen	Left	−22	17	8	165	2.48	0.0131			

FG, fusiform gyrus; HS/PT, Heschl sulcus/planum temporale; MTG, middle temporal gyrus; PCG, precentral gyrus; aSTG, anterior superior temporal gyrus; STS, superior temporal sulcus. *Average *t*-value and *P*-value for all voxels in a ROI. †Statistical tests used for ROI selection (corrected for cluster size at alpha = 5%).

2006). In contrast, posterior auditory brain areas, such as PT, have been implicated in integrating sensory with motor representations of speech (Buchsbaum *et al.*, 2005; Scott, 2005) and might therefore be involved in a 'how' pathway for audiovisual processing. It is interesting to consider the anatomical localization of posterior and anterior STS in this respect. Whereas posterior STS is anatomically closer to the dorsal processing pathways, anterior STS is closer to the ventral pathways (Sestieri *et al.*, 2006). Although speculative at this point, this observation underpins the existence of a neurofunctional distinction into a 'what' stream for cross-modal identity matching and a 'how' stream for matching sensory with motor representations.

A neural mechanism that can be put forward to account for the modulation of activity in STS and HS/PT by congruency is the neural integration of letters and speech sounds (van Atteveldt *et al.*, 2004). Congruent as opposed to incongruent letter–sound pairs represent overlearned audiovisual associations for which a strong neural association exists (Froyen *et al.*, 2008). Information about the identity of unisensory auditory and visual input has to be integrated in order to differentiate matching from non-matching letter–speech sound pairs and, hence, neural responses in STS and auditory cortex most likely reflect audiovisual integration.

Interestingly, as letters and speech sounds are not naturally linked before learning to read and their mapping has to be established through intensive training, the question rises which neural mechanisms support the perceptional shift from unfamiliar to familiar audiovisual objects for congruent letter–speech sound pairs. Hein *et al.* (2007) have investigated the relative influence of audiovisual object familiarity and congruency in recruiting cortical areas involved in audiovisual integration. Comparable to the present effects of congruency under low noise in STS and aSTG, the authors found posterior STS and STG to respond only for highly familiar and semantically congruent audiovisual stimuli. Other brain regions, such as inferior frontal cortex, in contrast responded to the learning of new audiovisual associations (Hein *et al.*, 2007). Similarly to the functional specialization for familiar objects that has been proposed for the visual system (Gauthier *et al.*, 1999), familiarity or frequency of exposure might also guide the functional specialization of neural systems for multisensory integration.

Finally, increased activation for congruent vs. incongruent letter–speech sounds was found in the FG (see Fig. 5), in close proximity to areas specialized for the processing of letter strings (Cohen *et al.*, 2002; McCandliss *et al.*, 2003; Cohen & Dehaene, 2004; Flowers *et al.*, 2004) or words (Kronbichler *et al.*, 2004). Modulation of FG by congruency has been previously reported in studies using audiovisual speech (Calvert *et al.*, 2000; Bushara *et al.*, 2003; Macaluso *et al.*, 2004). Interesting in this respect is the finding that the visual processing region typically described as FG might contain a lateral subregion, the lateral inferior multisensory area that essentially possesses multisensory response properties (Cohen & Dehaene, 2004). Thus far, neurofunctional models on the integration of letters and speech sounds do not incorporate a role for visual cortical regions in letter–sound integration (van Atteveldt *et al.*, 2004). The current findings, however, indicate that under certain task conditions effects of congruency are to be observed not only in the auditory cortex but also in the extrastriate visual cortex.

Behavioral relevance of letter–speech sound integration

The second main finding of the present study was that neural responses in the superior temporal cortex were similarly influenced by letter–sound congruency and visual noise as were behavioral responses to speech sounds. That is, the speed and accuracy of speech sound identification critically depended on the quality of the visual stimulus (congruency-by-noise interaction) where facilitation/inhibition effects were reduced as a function of letter quality. This finding suggests two things: first, it indicates that in STS/MTG basic stimulus features and stimulus content interact to mediate the cross-modal visual influences on speech identification, whereas earlier processing regions (HS/PT, FG) are influenced only by stimulus content. Previous findings have demonstrated a sensitivity of the left STS also for other binding factors such as temporal synchrony (Calvert *et al.*, 2000; Bushara *et al.*, 2003; Macaluso *et al.*, 2004) and various linguistic and non-linguistic audiovisual inputs (Calvert *et al.*, 2000; Beauchamp *et al.*, 2004), supporting a general integrative function for STS. Moreover, non-human electrophysiological work has indicated that STS, unlike PT for example, possesses an audiovisual response profile with convergent inputs from visual auditory and somatosensory cortex, which makes STS a likely candidate for early integration of AV inputs (Schroeder & Foxe, 2002).

Second and more importantly, demonstrating analogous effects of visual noise on neural responses to speech sounds in STS/MTG and

on the behavioral identification of speech sounds provides evidence for a role of STS/MTG in the cross-modal facilitation of speech perception. The present data therefore provide a critical extension to previous studies on audiovisual integration using either passive perception tasks (Calvert *et al.*, 2000; Olson *et al.*, 2002; van Atteveldt *et al.*, 2004, 2007) or indirect performance measures (Callan *et al.*, 2003; Macaluso *et al.*, 2004).

Summary and conclusion

We have described the results of a two-experiment behavioral and fMRI study, in which the influences of visual letters on speech sound identification were investigated. Our finding that letter–speech sound congruency modulated cortical responses to speech sounds in heteromodal superior temporal cortex as well as auditory and visual association cortex indicates automatic influences of visual letters on speech sound processing. More generally, these data support the existence of a strong mapping between orthography and phonology in literate adults. Moreover, the finding that adding visual noise to the letter stimulus led to a reduction of an effect of congruency, with remarkably similar neurofunctional (in STS/MTG) and behavioral response patterns (congruency-by-noise interaction), provides evidence for a functional involvement of heteromodal superior temporal cortex in the facilitation/inhibition of speech identification.

The present findings thus substantiate the hypothesis of an automatic neural mechanism for the integration of letter and speech sounds in literate adults that might be critical for associating spoken and written language. The acquisition and automatization of letter–speech sound associations is an important milestone in learning to read (Ehri, 2005). Failure to acquire sufficiently automatized letter–speech sound associations has been linked to dyslexia (Vellutino *et al.*, 2004). Therefore, knowledge on the automaticity and behavioral relevance of letter–speech sound integration in skilled readers adds to a basic framework for investigating the neural underpinnings of normal and abnormal reading development.

Acknowledgement

This research was partly supported by the Dutch Health Care Insurance Board (CVZ 608/001/2005 to L.B.).

Abbreviations

AVcon/AVinc, audiovisual congruent/incongruent; FG, fusiform gyrus; fMRI, functional magnetic resonance imaging; GLM, general linear model; HS/PT, Heschl sulcus/planum temporale; MTG, middle temporal gyrus; OT, occipito-temporal cortex; ROI, region of interest; STS/STG, superior temporal sulcus/gyrus.

References

van Atteveldt, N., Formisano, E., Goebel, R. & Blomert, L. (2004) Integration of letters and speech sounds in the human brain. *Neuron*, **43**, 271–282.

van Atteveldt, N.M., Formisano, E., Blomert, L. & Goebel, R. (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex*, **17**, 962–974.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H. & Martin, A. (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat. Neurosci.*, **7**, 1190–1192.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M. & Cox, R.W. (1996) Function of the left planum temporale in auditory and linguistic processing. *Brain*, **119**(Pt 4), 1239–1247.

Booth, J.R., Cho, S., Burman, D.D. & Bitan, T. (2007) Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. *Dev. Sci.*, **10**, 441–451.

Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B. & Mesulam, M.M. (2002) Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage*, **16**, 7–22.

Borowsky, R., Owen, W.J. & Fonos, N. (1999) Reading speech and hearing print: constraining models of visual word recognition by exploring connections with speech perception. *Can. J. Exp. Psychol.*, **53**, 294–305.

Buchsbaum, B.R., Olsen, R.K., Koch, P.F., Kohn, P., Kippenhan, J.S. & Berman, K.F. (2005) Reading, hearing, and the planum temporale. *Neuroimage*, **24**, 444–454.

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

Busse, L., Roberts, K.C., Crist, R.E., Weissman, D.H. & Woldorff, M.G. (2005) The spread of attention across modalities and space in a multisensory object. *Proc. Natl Acad. Sci. USA*, **102**, 18751–18756.

Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C. & Vatikiotis-Bateson, E. (2003) Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport*, **14**, 2213–2218.

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

Cohen, L. & Dehaene, S. (2004) Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, **22**, 466–476.

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S. & Dehaene, S. (2002) Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, **125**, 1054–1069.

Dijkstra, T., Schreuder, R. & Frauenfelder, U.H. (1989) Grapheme contex effects on phonemic processing. *Lang. Speech*, **32**, 98–108.

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

Ehri, L.C. (2005) Development of sight word reading:phases and findings. In Snowling, M.J. & Hulme, C. (Eds), *The Science of Reading: A Handbook*. Blackwell Publishing, Oxford, pp. 135–145.

Fiebach, C.J., Friederici, A.D., Muller, K. & von Cramon, D.Y. (2002) fMRI evidence for dual routes to the mental lexicon in visual word recognition. *J. Cogn. Neurosci.*, **14**, 11–23.

Fiez, J.A., Balota, D.A., Raichle, M.E. & Petersen, S.E. (1999) Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron*, **24**, 205–218.

Flowers, D.L., Jones, K., Noble, K., VanMeter, J., Zeffiro, T.A., Wood, F.B. & Eden, G.F. (2004) Attention to single letters activates left extrastriate cortex. *Neuroimage*, **21**, 829–839.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A. & Noll, D.C. (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.*, **33**, 636–647.

Frost, R. & Katz, L. (1989) Orthographic depth and the interaction of visual and auditory processing in word recognition. *Mem. Cognit.*, **17**, 302–310.

Froyen, D., van Atteveldt, N., Bonte, M. & Blomert, L. (2008) Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci. Lett.*, **430**, 23–28.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P. & Gore, J.C. (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.*, **2**, 568–573.

Goebel, R., Esposito, F. & Formisano, E. (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.*, **27**, 392–401.

Hasegawa, T., Matsuki, K.-I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y. & Sadato, N. (2004) Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cogn. Brain Res.*, **20**, 510–518.

Hashimoto, R. & Sakai, K.L. (2004) Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, **42**, 311–322.

Hein, G., Doehrmann, O., Muller, N.G., Kaiser, J., Muckli, L. & Naumer, M.J. (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.*, **27**, 7881–7887.

Hwang, J.H., Wu, C.W., Chen, J.H. & Liu, T.C. (2006) The effects of masking on the activation of auditory-associated cortex during speech listening in white noise. *Acta Otolaryngol.*, **126**, 916–920.

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

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W. & Ladurner, G. (2004) The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage*, **21**, 946–953.

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

Macaluso, E., George, N., Dolan, R., Spence, C. & Driver, J. (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage*, **21**, 725–732.

McCandliss, B.D., Cohen, L. & Dehaene, S. (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.*, **7**, 293–299.

McDonald, J.J., Teder-Salejarvi, W.A., Di Russo, F. & Hillyard, S.A. (2003) Neural substrates of perceptual enhancement by cross-modal spatial attention. *J. Cogn. Neurosci.*, **15**, 10–19.

Nakada, T., Fujii, Y., Yoneoka, Y. & Kwee, I.L. (2001) Planum temporale: where spoken and written language meet. *Eur. Neurol.*, **46**, 121–125.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C. & Rauschecker, J.P. (2006) Vowel sound extraction in anterior superior temporal cortex. *Hum. Brain Mapp.*, **27**, 562–571.

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

Perre, L. & Ziegler, J.C. (2008) On-line activation of orthography in spoken word recognition. *Brain Res.*, **1188**, 132–138.

Proverbio, A.M., Vecchi, L. & Zani, A. (2004) From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *J. Cogn. Neurosci.*, **16**, 301–317.

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

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

Scott, S.K. (2005) Auditory processing – speech, space and auditory objects. *Curr. Opin. Neurobiol.*, **15**, 197–201.

Scott, S.K. & Johnsrude, I.S. (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci.*, **26**, 100–107.

Scott, S.K., Blank, C.C., Rosen, S. & Wise, R.J. (2000) Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, **123**(Pt 12), 2400–2406.

Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J.G., Bilecen, D., Tedeschi, G., Scheffler, K. & Di Salle, F. (2002) Spatiotemporal pattern of neural processing in the human auditory cortex. *Science*, **297**, 1706–1708.

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

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Olivetti Belardinelli, M. & Romani, G.L. (2006) "What" versus "where" in the audiovisual domain: an fMRI study. *Neuroimage*, **33**, 672–680.

Simos, P.G., Breier, J.I., Fletcher, J.M., Foorman, B.R., Castillo, E.M. & Papanicolaou, A.C. (2002) Brain mechanisms for reading words and pseudowords: an integrated approach. *Cereb. Cortex*, **12**, 297–305.

Stanford, T.R., Quessy, S. & Stein, B.E. (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.*, **25**, 6499–6508.

Tagamets, M.A., Novick, J.M., Chalmers, M.L. & Friedman, R.B. (2000) A parametric approach to orthographic processing in the brain: an fMRI study. *J. Cogn. Neurosci.*, **12**, 281–297.

Talairach, J. & Tournoux, P. (1988) *Co-Planar Stereotactic Atlas of the Human Brain*. Thieme, Stuttgart.

Vellutino, F.R., Fletcher, J.M., Snowling, M.J. & Scanlon, D.M. (2004) Specific reading disability (dyslexia): what have we learned in the past four decades? *J. Child Psychol. Psychiatry*, **45**, 2–40.

Wilson, S.M., Saygin, A.P., Sereno, M.I. & Iacoboni, M. (2004) Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.*, **7**, 701–702.

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