

# Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity

Citation for published version (APA):

van Atteveldt, N. M., Roebroek, A. F., & Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hearing Research*, 258(1-2), 152-164. <https://doi.org/10.1016/j.heares.2009.05.007>

## Document status and date:

Published: 01/01/2009

## DOI:

[10.1016/j.heares.2009.05.007](https://doi.org/10.1016/j.heares.2009.05.007)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

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

[Link to publication](#)

## General rights

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

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

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

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

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

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



## Research Papers

## Interaction of speech and script in human auditory cortex: Insights from neuro-imaging and effective connectivity

Nienke van Atteveldt<sup>a,b,\*</sup>, Alard Roebroek<sup>a</sup>, Rainer Goebel<sup>a,c</sup><sup>a</sup> Dept. of Cognitive Neuroscience, Faculty of Psychology & Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands<sup>b</sup> Dept. of Psychiatry, Division of Child and Adolescent Psychiatry, Columbia University College of Physicians and Surgeons & New York State Psychiatric Institute, 1051 Riverside Drive, New York, NY 10032, USA<sup>c</sup> Netherlands Institute for Neuroscience, Meibergdreef 47, 1105BA Amsterdam, The Netherlands

## ARTICLE INFO

## Article history:

Received 11 December 2008

Received in revised form 18 May 2009

Accepted 20 May 2009

Available online 13 June 2009

## Keywords:

fMRI

Auditory cortex

Audiovisual

Speech

Script

Multisensory integration

Granger causality mapping

## ABSTRACT

In addition to visual information from the face of the speaker, a less natural, but nowadays extremely important visual component of speech is its representation in script. In this review, neuro-imaging studies are examined which were aimed to understand how speech and script are associated in the adult “literate” brain. The reviewed studies focused on the role of different stimulus and task factors and effective connectivity between different brain regions. The studies will be summarized in a neural mechanism for the integration of speech and script that can serve as a basis for future studies addressing (the failure of) literacy acquisition. In this proposed mechanism, speech sound processing in auditory cortex is modulated by co-presented visual letters, depending on the congruency of the letter–sound pairs. Other factors of influence are temporal correspondence, input quality and task instruction. We present results showing that the modulation of auditory cortex is most likely mediated by feedback from heteromodal areas in the superior temporal cortex, but direct influences from visual cortex are not excluded. The influence of script on speech sound processing occurs automatically and shows extended development during reading acquisition. This review concludes with suggestions to answer currently still open questions to get closer to understanding the neural basis of normal and impaired literacy.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

Speech is a multimodal phenomenon. Most naturally, speech is highly intertwined with visual information from the speaker's face. This includes, but is not confined to, lip movements (Campanella

and Belin, 2007). Many neuro-imaging studies have shown that visual information from the speaker can by itself activate the auditory cortex (Campbell, 2008; Molholm and Foxe, 2005), or modulate speech processing during stages previously assumed to be unimodal auditory (Amedi et al., 2005; Calvert et al., 1999). This observation is not unique to speech processing, as visual influences in “auditory” cortex have been observed across species and for many types of stimuli (Kayser and Logothetis, 2007; Schroeder and Foxe, 2005).

In addition to visual information from the (face of the) speaker, a less natural, but nowadays extremely important visual component of speech is its representation in written language, or *script*. Since spoken language is our primary language system, learning to read essentially comes down to understanding how a script represents the speech system. In alphabetic scripts, such as English or Dutch, letters and speech sounds (or *phonemes*) are the basic elements of correspondence between spoken and written language. Investigations of how letters and speech sounds are associated in the brain therefore increase our understanding of the neural basis of literacy, the relevance of which is emphasized by the constantly increasing demands for literacy in today's knowledge-based society (National Reading Council, 1998).

**Abbreviations:** A1, primary auditory cortex; aSTP, anterior superior temporal plane; BOLD, Blood Oxygen Level Dependent; CSEM, covariance structural equation modeling; DCM, dynamic causal modeling; EEG, electro-encephalography; er-fMRI, event-related fMRI; ERP, event-related potential; FDR, False Discovery Rate; fMRI, functional magnetic resonance imaging; fMR-A, fMR adaption; GCM, Granger causality mapping; GLM, General Linear Model; HS, Heschl's sulcus; ICA, independent component analysis; MANOVA, multivariate analysis of variance; MEG, magneto-encephalography; MMN, mismatch negativity; MSI, multisensory integration; VPA, multi-voxel pattern analysis; PAC, primary auditory cortex.; PPI, psycho-physiological interactions; PT, planum temporale; PTa, anterior PT; PTP, posterior PT; ROI, region-of-interest; SOA, stimulus onset asynchrony; STS, superior temporal sulcus; STG, superior temporal gyrus; SVM, Support Vector Machine; V1, primary visual cortex

\* Corresponding author. Address: Dept. of Psychiatry, Division of Child and Adolescent Psychiatry, Columbia University College of Physicians and Surgeons, 1051 Riverside Drive, New York, NY 10032, USA. Tel.: +1 212 543 5995.

E-mail addresses: [n.vanatteveldt@maastrichtuniversity.nl](mailto:n.vanatteveldt@maastrichtuniversity.nl) (N. van Atteveldt), [a.roebroek@maastrichtuniversity.nl](mailto:a.roebroek@maastrichtuniversity.nl) (A. Roebroek), [r.goebel@maastrichtuniversity.nl](mailto:r.goebel@maastrichtuniversity.nl) (R. Goebel).

In this review, neuro-imaging investigations of how speech and script are associated in the adult “literate” brain will be closely examined to formulate a neural mechanism that can serve as a basis for future studies addressing (the failure of) literacy acquisition. In the past few years, it has become apparent that the auditory cortex plays an important role in the integration of speech and script, which may not be surprising given the fact that speech is primary to the learning and use of a script. To provide a theoretical framework for this view, this review will start by giving a more elaborate background on the relation between spoken and written language. Next, empirical evidence as to where, when and how speech and script information interact in the brain will be reviewed. For the “how” question, of special interest are (1) the role of different stimulus and task factors and (2) effective connectivity within the network, i.e., the question of how different brain regions cooperate to successfully connect speech with its written representation.

The different stimulus manipulations that have been investigated include semantic congruency (Van Atteveldt et al., 2004), temporal correspondence (Van Atteveldt et al., 2007a), input quality (Blau et al., 2008a) and stimulus repetition (van Atteveldt et al., 2008). To investigate how different top-down task demands influence the perceptual integration processes, we compared passive and active matching (Van Atteveldt et al., 2007b) and the effects of explicit versus implicit integration processes during task performance. These factors will be considered one by one, before they will be merged into a neural mechanism for letter–sound integration, in which preliminary effective connectivity results will also be included. Finally, suggestions for future studies to answer currently still open questions will be discussed, to get closer to understanding the neural basis of normal and impaired literacy.

## 2. The relation between speech and script

Reading has been characterized as being “parasitic on speech” (Mattingly, 1972), and learning to read as “an unnatural act” (Gough and Hillinger, 1980). These statements refer to the difference in naturalness between spoken and written language. Spoken language precedes written language, both in phylogenetic and ontogenetic development. In the time-course of evolution, spoken language is as old as the human species, while writing systems are cultural inventions that only exist since a few thousands years, and are of common use only since the last century. As a consequence, brain mechanisms for spoken language are a product of biological evolution, whereas the artificial nature and relatively short history of writing systems make the existence of a naturally adapted specialized brain mechanism for written language unlikely (Gleitman and Rozin, 1977; Liberman, 1992). At the same time, literacy has become extremely important since its short existence. In many Western countries, 100% of the population is expected to be literate nowadays (National Reading Council, 1998). Considering the artificial nature of written language, meeting this expectation logically poses a serious challenge during individual cognitive development.

In analogy with evolutionary development, speech is also earlier and more natural during individual development. Whereas spoken language is typically acquired automatically and without effort, literacy takes a long period of instruction (Froyen et al., 2009), and is usually learned at an older age (Eden and Moats, 2002), if learned at all. As already mentioned, the essential step in becoming literate is to discover *how* a writing system encodes the already acquired spoken language system. In alphabetic scripts, this is called the *alphabetic principle* (Byrne et al., 1996), which includes learning the letter–sound correspondences of a language. Different developmental theories are in agreement that, although other factors also play a role, this is the crucial step in literacy

acquisition (Ehri, 2005). Only after mastering all letter–sound mappings, all words can be decoded systematically. Skilled reading most probably still involves automatic activation of phonological representations (National Reading Council, 1998; Perfetti and Sandak, 2000; Snowling and Hulme, 2005). This indicates that after successful acquisition of the alphabetic principle and many years of reading experience, letter–speech sound associations are over-learned paired associates in literate adults.

To sum up, in speech-based alphabetic scripts, the spoken language system is primary and is represented by script at the level of letters and speech sounds. To gain more insight in the neural basis of literacy skills, letter–speech sound associations are the most fundamental level to study. Therefore, although several neuro-imaging studies have studied letter-to-sound conversion indirectly or at the level of words (e.g., see Price and Mechelli, 2005), this review will focus on studies that provide direct insight into how letters and speech sounds interact in the brain.

## 3. Empirical evidence for interaction of speech and script in the auditory cortex

### 3.1. Functional magnetic resonance imaging: a short note on statistical criteria

The question of what is the best metric to infer multisensory integration (MSI) from functional magnetic resonance imaging (fMRI) data is a matter of ongoing debate (Beauchamp, 2005; Calvert, 2001; Driver and Noesselt, 2008; Laurienti et al., 2005). We favor the criterion that the bimodal fMRI-response, or Blood Oxygen Level Dependent (BOLD) response, should be increased relative to the most effective unimodal response, also termed the “max criterion” (Beauchamp, 2005). A similar criterion has been used to infer multisensory enhancement or depression on the single neuron level, using the number of evoked impulses to unimodal and bimodal stimulation (Stein and Meredith, 1993). The max criterion is more stringent than the also used “mean criterion” in which the bimodal response should be more than the mean of the unimodal responses. This criterion is too liberal especially when one of the unimodal responses is weak or absent, which reduces the mean in such a way that a bimodal response exceeds the mean even when weaker than the largest unimodal response (Beauchamp, 2005). Therefore, the mean criterion can be misleading, especially when examining low-level sensory regions such as the auditory cortex, since it would be classified as “multisensory” when visual responses are weak, even if the audiovisual response is weaker than the auditory response.

On the other hand, the max criterion is more liberal than the super-additivity metric, which states that the bimodal response should exceed the sum of the unimodal responses. Although claimed to be the only metric that can infer MSI from fMRI data with high certainty (i.e., without false positives, Calvert, 2001), it appears to be too strict, likely leading to many false negatives. This is because the high variability in operational modes of single multisensory neurons (Perrault et al., 2005) in combination with the indirect nature, limited dynamic range and coarse spatial resolution of fMRI makes a super-additive BOLD response rather unlikely (Laurienti et al., 2005). All in all, the max criterion seems to be the best compromise, as is supported by a direct comparison of the different analytic metrics (Beauchamp, 2005).

Because all different metrics for comparing bimodal to unimodal responses have limitations, an interesting additional approach is to manipulate the (e.g., identity) congruency of the different inputs (e.g., visual and auditory). In this type of analysis, two bimodal conditions are contrasted with each other (congruent vs. incongruent), which eliminates the unimodal component and its accompanying

complications from the metric. Since a distinction between congruent and incongruent cross-modal stimulus pairs can not be established unless the unimodal inputs have been integrated successfully, the congruency contrast can be used as a supplemental criterion for multisensory integration. An additional advantage of using the congruency contrast is that it only has two levels, which facilitates manipulation of different factors (e.g., temporal or spatial relation between the cross-modal inputs) within the same design and directly search for interactions using multi-factorial analyses (Van Atteveldt et al., 2007a). For example, interactions between temporal or spatial relation and identity congruency directly address the impact of these factors on multisensory integration.

### 3.2. Visual letters modulate speech sound processing in low-level auditory cortex

In the first of a series of fMRI studies in our group, we aimed to elucidate the network of brain regions involved in the processing, convergence and integration of letters and speech sounds in literate adults (Van Atteveldt et al., 2004). To achieve this aim, we presented our subjects with single letters (unimodal visual), speech sounds (unimodal auditory), and their combinations (bimodal/audiovocal) during fMRI scanning. Moreover, we manipulated the identity congruency of the bimodal stimuli: congruent (e.g., letter “a” combined with speech sound /a/) and incongruent (e.g., letter “a” with speech sound /o/) pairs were presented in different blocks. The unimodal conditions were included to demonstrate processing and convergence of letters and speech sounds, by testing which areas responded specifically to only letters or only speech sounds (modality-specific processing), and which responded to both (convergence or heteromodal processing). The bimodal conditions were included to assess which areas were involved in multisensory integration, by contrasting bimodal to unimodal responses (using the max criterion), and bimodal congruent and incongruent conditions to each other (congruency contrast).

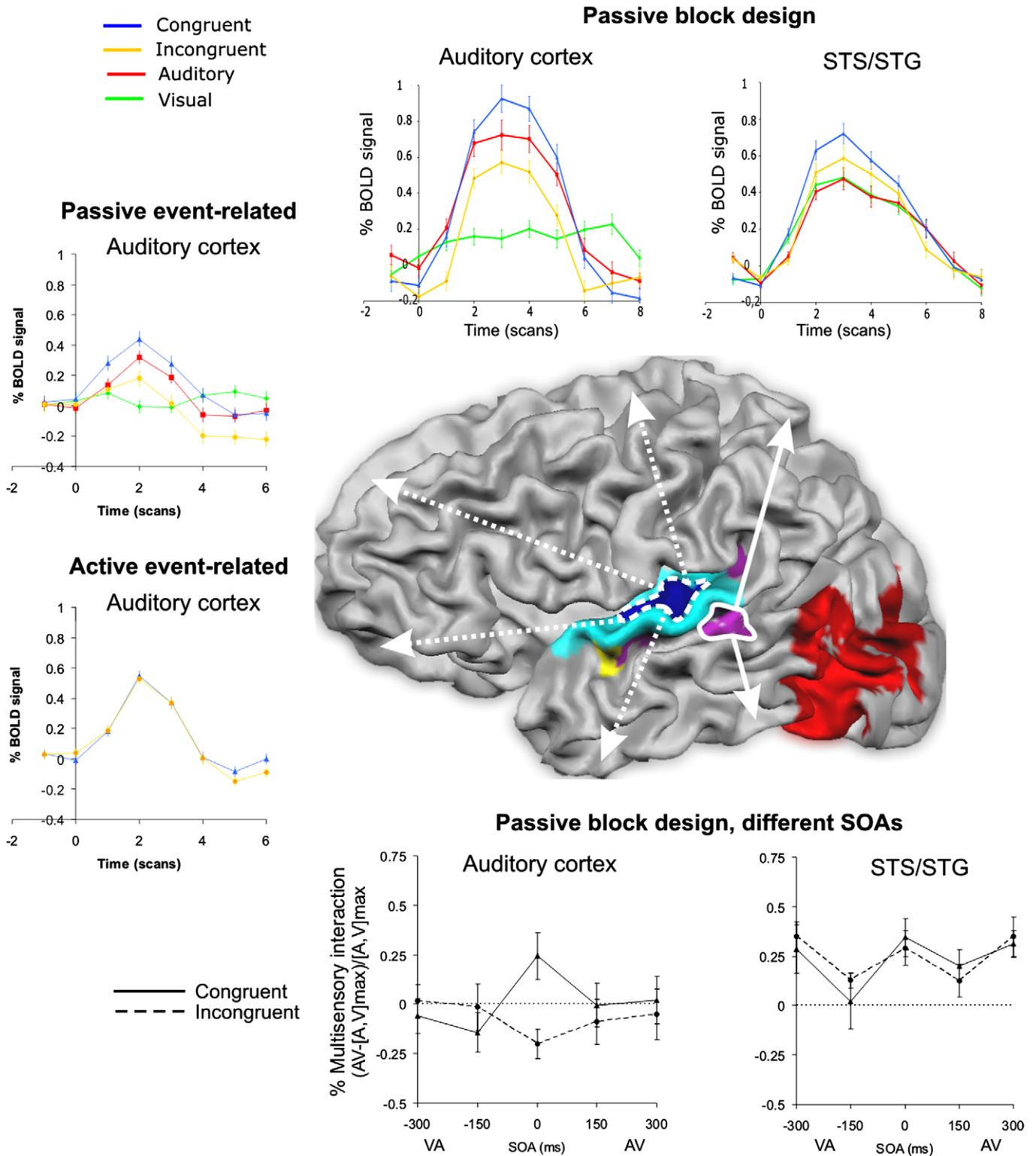
The key findings from this fMRI study were that: (1) Regions of the superior temporal sulcus and gyrus (STS/STG) showed response enhancement for both bimodal conditions, and a converging or heteromodal response pattern to the unimodal conditions; and (2) Compared to the fMRI response to speech sounds alone, low-level (“unimodal”) auditory cortex regions showed an increased response to congruent and a suppressed response to incongruent letter–sound pairs (congruency effect), in absence of a significant response to letters alone. These two findings are summarized in Fig. 1 (graphs in top row: “passive block design”). The congruency effect was located on Heschl’s sulcus (HS), extending to the planum temporale (PT), probably corresponding to auditory belt cortex as described in non-human primates (Kaas and Hackett, 2000), reflecting the second stage of processing (after processing in the primary core). The absence of a clear unimodal visual response in PT/HS indicated that the effect of the letters on speech sound processing in this area reflects a modulatory feedback effect with likely origin in the STS/STG, as was also suggested for audiovisual speech (Calvert, 2001). However, an interesting alternative interpretation of the visual modulation of PT/HS activity is by direct anatomical connections between visual and auditory cortex that have been demonstrated in animals (Budinger et al., 2006; Falchier et al., 2002; Rockland and Ojima, 2003). This should be explored in more detail in future studies, for instance using fMRI effective connectivity measures (see Section 4.3), but complementary methods providing precise timing information might be necessary for conclusive answers (see Section 6.2).

The exact processing stage in which letters modulated speech sound processing (secondary auditory cortex) was estimated on basis of individual anatomical landmarks on flattened cortical

sheets that provided detailed anatomical information (figure 4 in Van Atteveldt et al., 2004). As also clear from this figure, the anatomy of Heschl’s region is highly variable across subjects (Morosan et al., 2001). Therefore, it is of great importance to examine the anatomical location of visual influences in the auditory cortex on an individual basis, as was done very systematically in a study on lip reading (Pekkola et al., 2005). However, as also indicated by Pekkola and colleagues, the exact topography of functional fields within auditory cortex also varies, and is inconsistently related to the macro-anatomical landmarks available from MRI images (Morosan et al., 2001). This makes it impossible to delineate primary from secondary auditory cortex with certainty based on anatomical MRI’s only. More ideally, processing stage should be defined functionally using tonotopic mapping, in analogy to retinotopic mapping which is used in the visual cortex to delineate early visual areas (Serenio et al., 1995). Using high-resolution fMRI, Formisano et al. (2003) convincingly demonstrated that the human primary auditory cortex (PAC) consists of two mirror-reversed tonotopic maps, i.e., exhibiting a low-to-high or high-to-low gradient of frequency tuning. This corresponds with what has been shown using invasive electrophysiology in the macaque PAC (Kaas and Hackett, 2000). This knowledge can be applied in studies investigating different aspects of sound processing, to verify whether effects are located in the PAC or not, and in which subdivision (see e.g., Riecke et al., 2007). Analogously, it will be very interesting to use such best-frequency mapping techniques in human multisensory studies, to functionally rather than only anatomically verify the exact processing stage of visual influences (see also Section 6.1.1). In a high-resolution fMRI study on visual influences in the macaque auditory cortex, Kayser and colleagues mapped auditory cortex fields (Kayser et al., 2007) using a technique that combines differential frequency and bandwidth preferences (Petkov et al., 2006). In their study, Kayser and colleagues were able to demonstrate very accurately that strongest visual influences occurred in auditory (para-) belt, but extended to (caudal) primary cortex. The challenge now lies in combining a similar sophisticated form of functional mapping with multisensory experiments in humans as well.

### 3.3. Influence of input timing

In the study that followed, we manipulated temporal correspondence between the letters and speech sounds in the bimodal conditions, in addition to the identity congruency manipulation (Van Atteveldt et al., 2007a). As discussed in Section 3.1, this is one of the advantages of the congruency analysis approach. In both bimodal congruent and incongruent conditions, we now used five different stimulus onset asynchronies (SOA) between the letter and speech sound stimuli: –300, –150, 0, 150 and 300 ms (negative SOA’s indicate that visual precedes auditory). This balanced 2-factorial ( $2 \times 5$ ) design allowed a voxel-by-voxel search for interactions between congruency and SOA, in other words, voxels that show a different effect of congruency for different input timings. Interestingly, the interaction analysis revealed auditory cortex areas corresponding to the PT/HS regions that showed the congruency effect in the previous study. As summarized in Fig. 1 (“Auditory cortex” graph in bottom row: “Passive block design, different SOA’s”), the PT again showed a strong congruency effect, but only for synchronously presented visual and auditory inputs. The multisensory enhancement for congruent, as well as the suppression for incongruent letter–sound pairs rapidly diminished with increasing SOA. In addition, a bilateral, more anterior region on the superior temporal plane (aSTP) showed a significant interaction between SOA and congruency (not shown here, see Figures 3–5 in Van Atteveldt et al., 2007a). These aSTP regions exhibited the congruency effect also only for synchronous presentation,



**Fig. 1.** Summary of response profiles in auditory cortex and STS/STG as influenced by modality, congruency and temporal asynchrony of letter–sound pairs, under different task conditions. (Center) The brain depicts the reconstructed left hemisphere on which different statistical maps are projected: congruency contrast in dark blue, max criterion in violet. The dark blue region encircled by the dotted line reflects the auditory cortex; the violet region encircled by the closed line reflects heteromodal STS/STG. The figure is adapted from van Atteveldt et al. (2004), but the maps are representative also for those obtained in the other fMRI studies. (Graphs) The top row (“passive block design”) displays the averaged % BOLD response time-course in the auditory cortex (left graph) and STS/STG (right graph) in the first fMRI study in which stimuli were blocked, audiovisual stimuli synchronous, and viewing/listening was passive. The bottom row (“passive block design, different SOAs”) shows the multisensory interaction index as a function of SOA (VA, visual precedes; AV, auditory precedes). The index reflects the audiovisual (AV) response as percentage of the most effective unimodal response (auditory/A or visual/V) and was calculated by:  $((AV - [A,V]_{max})/[A,V]_{max}) \times 100\%$ . The left panel (“passive/active event-related”) shows the averaged % BOLD response time-course in the auditory cortex using event-related fMRI (random stimulus presentation) during passive viewing/listening (upper graph) and active matching (lower graph) of letter–sound pairs.

although the multisensory interaction values seemed to fluctuate more as a function of SOA.

Because the SOA by congruency interaction analysis did not reveal voxels in the heteromodal STS/STG, we applied the max criterion at all five SOA's to examine the effect of input timing on MSI independent of congruency. This analysis revealed response enhancement for both congruent and incongruent letter–sound pairs relative to the most effective unimodal response in the STS/STG, roughly for all SOA's (Fig. 1, “STS/STG” graph in bottom row: “Passive block design, different SOA's”). Interestingly, as shown in Fig. 1, the strength of response enhancement seems to fluctuate as a function of SOA: some SOAs seem to be “better” than others for MSI. This hints at a recently revealed and potentially very interesting mechanism in which cross-modal inputs can influence each other by resetting the phase of ongoing oscillations (Ghazanfar and Chandrasekaran, 2007; Lakatos et al., 2007). Ongoing oscillations are fluctuations of local neuronal ensembles between high and low excitability states (Schroeder et al., 2008). Lakatos and colleagues found in the macaque auditory cortex that the phase of ongoing oscillations could be reset by somatosensory input in such a way that the processing of subsequent auditory inputs was enhanced or suppressed, depending on whether they arrived during an excitatory (“best”) or inhibitory (“worst”) phase of the oscillatory cycle of a certain frequency band. Enhancement peaks corresponded to gamma, theta and delta periods. Moreover, they verified phase resetting by comparing phase distribution of different frequency bands before and after somatosensory stimulation. This revealed that after somatosensory stimulation, the phase distribution of oscillations in exactly the gamma, theta and delta bands were concentrated relative to prestimulus, when they were random. The SOA fluctuation that we observed in STS seems consistent with the delta band period and might therefore point to a similar mechanism. However, note that we sampled only a limited set of SOAs, and that the auditory stimulus preceded in half of them.

In relation to the previous fMRI study described in Section 3.2, we interpreted the SOA effects as follows: letters and speech sounds are integrated in STS over a wide temporal range (though with better and worse SOAs), but feedback to the auditory cortex, in this study PT and aSTP, is only provided when the stimuli are in temporal synchrony. This is supported by another fMRI study comparing synchronous with asynchronous audiovisual patterns (Noesselt et al., 2007). Their effective connectivity analysis indicated feedback from STS to auditory cortex only for the coincident stimuli.

### 3.4. Impact of different task demands

So far, our studies only revealed insights in letter–sound integration in a passive setting, as no explicit top–down task instructions were imposed during viewing and listening. We therefore believe that the results described in Sections 3.2 and 3.3 reflect an automatic, bottom–up mechanism of integration, where only stimulus properties (identity congruency, temporal correspondence) define which inputs will be integrated (and how). To investigate how different top–down task demands influence the integration process, we included an active condition in a subsequent study (Van Atteveldt et al., 2007b). In this study, we first switched from blocked to random stimulus presentation during passive perception, which served two goals. Firstly, since the use of a behavioral task necessitates random stimulus presentation and thereby event-related fMRI (er-fMRI) analysis, we wanted to replicate the previously demonstrated multisensory effects using er-fMRI. The second goal was to exclude the possibility that blocked presentation of congruent and incongruent stimuli evoked different attention conditions that could (partly) be responsible for

the demonstrated congruency effects. Both goals were achieved: er-fMRI analysis of unimodal and bimodal congruent and incongruent stimuli almost perfectly replicated the results from the blocked stimulus presentation in the auditory cortex (Fig. 1, “passive event-related” graph in left column). This confirms that the congruency effect we observed using blocked stimulus presentation is not (exclusively) attributable to attention differences in the congruent and incongruent blocks. Note that the different shape and amplitude of the event-related BOLD response are as expected from the shorter stimulus durations.

In the second part of the study, only bimodal (congruent and incongruent) stimuli were randomly presented, and subjects were instructed to indicate by a button press whether or not the letter matched the speech sound (“yes” or “no”). The stimulus presentation and scanning design was identical to the passive event-related experiment. Interestingly, during active matching, the congruency effect was no longer observed in the auditory cortex or in any other region (Fig. 1, “active event-related” graph in left column), at any of the tested SOA's (0 and  $\pm 300$  ms). When both event-related graphs are compared, the absence of the congruency effect seems to be best explained by an increased response to incongruent stimuli during explicit matching. An additional whole-brain analysis revealed that several frontal and parietal areas were stronger activated during incongruent relative to congruent stimuli. These results demonstrate that multisensory responses to letter–sound pairs in the auditory cortex heavily depend on task demands. Whereas the auditory cortex is modulated by congruency and temporal synchrony in absence of task instructions, this modulation is overruled during a matching task for which congruency information is explicitly needed for performance. The increased activity for incongruent versus congruent letter–sound pairs in frontal and parietal regions indicate a possible source of this task-related top–down effect.

An intriguing question that followed from this study is whether congruency information *will* influence auditory cortex when it is *not* an explicit aspect of the task. This question was addressed in another fMRI study in which subjects performed a unimodal auditory speech sound discrimination task (“did you hear /a/ or /e/?”, Blau et al., 2008a). The speech sounds were presented unimodally or combined with congruent or incongruent letters; however, the letters were completely irrelevant for task performance. The nature of the task makes integration of the letters and speech sounds, if it occurs, implicit since it is not needed for solving the task. In the auditory cortex, a similar congruency effect was observed as during passive integration, indicating that implicit integration during an active task does not overrule this response. In sum, different task demands have a different impact on the neural mechanism of letter–speech sound integration, and especially the explicitness of the integration process seems to influence “default” responses in the auditory cortex as observed during passive viewing and listening conditions. This implies that one should consider carefully what kind of task would be best suited for a specific purpose.

### 3.5. Input quality

Combination of inputs from different senses is especially helpful in circumstances where the separate inputs do not provide optimal information. Examples are the improvement of speech perception by lip reading in noisy surroundings or the complementary auditory or somatosensory information provided when dusk falls. This phenomenon was exploited in the study of Blau and colleagues (2008a, see 3.4), where sub-optimal stimuli were used in both modalities: pink noise was added to the speech sounds, and the letters were degraded by overall contrast reduction and masking with an abstract reassembling of letter parts. In addition to the

results described in Section 3.4, the visual cortex (fusiform gyrus) also exhibited an effect of letter–sound congruency, which was remarkable for two reasons: (1) visual information was irrelevant for the task (speech sound discrimination) and (2) a congruency effect was not found in the visual cortex using undegraded letter and speech sound stimuli (e.g., van Atteveldt et al., 2004). The first reason emphasizes how automatic the integration of letters and speech sound occurs in literate adults, since the letters were irrelevant for performing the task. The second reason suggests that the quality of the input can influence the direction of feedback modulation. Apparently, when stimuli are sub-optimal, the visual cortex is also involved in the neural mechanism for letter–speech sound integration, possibly also by feedback from STS. The visual region affected by letter–sound congruency was located on the fusiform gyrus and corresponds to regions specialized for letter and word processing described by others (Cohen and Dehaene, 2004; Flowers et al., 2004; Polk et al., 2002).

### 3.6. Evidence from electrophysiology: when

Due to the limited temporal resolution of fMRI, it can normally not provide information on *when* multisensory effects occur except for the case when it refers to processing stage rather than time. Therefore, precise temporal information using electrophysiology is indispensable. Raij et al. (2000) investigated the time-course of letter–sound integration using magneto-encephalography (MEG), and demonstrated the time-course of a cascade of neural events that fits our fMRI findings surprisingly well. They found convergence (i.e., overlap) of activation to visual and auditory letter and control stimuli starting from 225 ms in lateral temporal cortex. Next, they compared audiovisual and unimodal responses to determine multisensory interaction. Different interaction strength for matching and nonmatching audiovisual letters, both stronger than control stimuli, was found later in time (380–540 ms). This latter effect is compatible with the congruency effect we observed in auditory cortex, and the late onset with our feedback interpretation.

Based on the insight from our fMRI studies that letter–sound congruency has a strong effect on auditory processing of speech sounds, we conducted event-related potential (ERP) studies using the mismatch negativity (MMN) to speech sound deviants combined with letters or not. An MMN is typically evoked when an infrequently presented sound (the deviant) deviates in certain aspects from a frequently presented sound (the standard), and is generally thought to reflect a pre-attentive auditory change detection mechanism (for reviews, see Picton et al., 2000; Schröger, 1998). We compared the MMN to deviant speech sounds when the standard speech sounds were presented alone versus when they were combined with matching letters. Moreover, we examined this relative MMN amplitude at different SOAs between the letters and speech sounds, in adults (Froyen et al., 2008) and in children of different reading levels (Froyen et al., 2009).

The results of the adult study support our fMRI finding that auditory processing is influenced by letter–sound congruency: the MMN to a deviant speech sound (peaks between 170 and 190 ms) was enhanced when the standard speech sounds were combined with matching letters, in other words, the deviant not only deviated from the speech sound, but also violated the letter–sound congruency. Interestingly, the MMN enhancement was strongest when letters and speech sounds were presented simultaneously and significantly weaker at 100 and 200 ms SOA, which is also consistent with our fMRI results. Furthermore, the enhanced MMN was not observed in beginner readers, only a later effect in the difference wave (around 650 ms). Advanced readers did exhibit an enhanced effect within the MMN time window, however, only at 200 ms SOA. These results suggest a protracted developmental course towards automated (i.e., adult-like) letter–sound associations.

In sum, the MMN results support our findings from fMRI that auditory processing of speech sounds can be influenced by letter–sound congruency and temporal asynchrony, and further support that the effects are automatic or pre-attentive. Moreover, they extend our previous findings with timing and developmental data, showing that the effects can occur as early as 180 ms after auditory stimulus onset (literate adults) but also that this early effect only occurs after a long period of reading acquisition. The latency of the MMN enhancement (180 ms) suggests a direct interaction between visual and auditory cortices, which is also supported by the increased power (2–10 Hz) within the first 250 ms after congruent grapheme–phoneme pairs in left auditory cortex reported by Herdman et al. (2006). However, it should be noted that the MMN does not simply reflect auditory processing but a change detection mechanism (Picton et al., 2000), and that in our studies, only the auditory stimulus deviated. The cross-modal context during the audiovisual standards might also be created by feedback from STS, as supported by the timing data from Raij et al. (2000). Further ERP (or MEG) investigations are needed that provide more decisive results on feedforward versus feedback influences of letters on speech sound processing. In addition, investigating effective connectivity between uni- and multisensory brain areas might shed light on the issue, as will be described next.

## 4. Effective connectivity during multisensory integration

### 4.1. Effective connectivity analysis of fMRI data: different approaches

Before discussing various recent fMRI-based connectivity investigations into multisensory integration, it is important to get some insight into the different methodological approaches that have been used. Rather than providing an extensive overview of all methods and their relative strengths and weaknesses, which is beyond the scope of this review, we focus on a few distinguishing characteristics of fMRI connectivity approaches that form useful contrasts in their relative benefits. First, there is the important distinction between functional connectivity, which amounts to investigating the correlation between measured time-courses of different brain areas (Friston et al., 1993b), and effective connectivity, defined as the influence one neuronal system exerts over another (Friston, 1994; Friston et al., 1993a). Popular fMRI functional connectivity approaches include principle component analysis (Friston et al., 1993b) and independent component analysis or ICA (McKeown et al., 1998). Whereas functional connectivity is ambiguous with respect to underlying directed interactions that generated the observed correlations, effective connectivity approaches attempt to resolve this ambiguity by defining explicit statistical models of directed neuronal interactions. Recently, functional connectivity approaches, particularly ICA, have been increasingly used to study brain networks during rest (Greicius et al., 2003; van de Ven et al., 2004), whereas effective connectivity approaches are almost exclusively used during task performance. Generally, an effective connectivity model consists of a structural or neuro-anatomical model, specifying *which* brain areas are thought to interact, and a mathematical model specifying *how* interactions between areas are expressed in measured signals (Büchel and Friston, 2000). The second important distinguishing dimension pertains to the employed structural model and lays out a continuum from approaches with a very strict model, containing few (i.e., 3–8) pre-selected regions and directed connections between them in a small directed graph model, via approaches with less restrictive structural models, allowing for many interacting regions, to even mapping interacting regions over the whole brain. Methods on the restrictive end of the continuum include covariance structural equation modeling (CSEM, Büchel

and Friston, 1997; McIntosh and Gonzalez-Lima, 1994) and dynamic causal modeling (DCM, Friston et al., 2003). It should be stressed that pre-specified structural models are highly useful in specifying and testing specific hypotheses concerning interactions within well-known systems of brain areas. However, an implicit problem in this type of approach is that misspecification of the models (e.g., by omission of an area that mediates or initiates interactions) can lead to erroneous conclusions. The approaches with more liberal structural models include that of Deshpande et al. (2008) which computes connectivity between a large set of 25 brain areas, and Granger causality mapping (GCM, Roebroeck et al., 2005) that maps all areas in the brain interacting with a pre-selected reference area. The method of psycho-physiological interactions (PPI, Friston et al., 1997) can also be used to map areas in the brain that play a driving or modulatory role within a partly specified structural model.

The third important distinguishing aspect of effective connectivity approaches pertains to their mathematical model, namely whether it is static or dynamic. A static mathematical model derives connectivity estimates only from statistical relations (e.g., covariation) of signals at the same timesample. Therefore, static effective connectivity methods, like CSEM and PPI derive the direction of influence exclusively from the prespecified structural model. Dynamic effective connectivity models, on the other hand, are able to use the dynamic interrelation of signals over time as the basis of their connectivity estimates and, to some degree, are able to use the arrow of time in deriving directions of interaction. Within the dynamic models a distinction can be made between deterministic and stochastic models. Deterministic signal models, such as DCM (Friston et al., 2003), have noise terms only at the observation level such that any dynamic signal variation is strictly determined by the input to the model. Therefore, the prespecified model input (often specified ‘predictor’ functions that form the design matrix for a General Linear Model or GLM analysis) fully determines the signal variations and effective connectivity that can be estimated. In contrast, stochastic models, such as multivariate autoregressive models (e.g., Roebroeck et al., 2005) model signal variations as modulation of input noise. This leaves the possibility to capture unforeseen signal variation. All models of fMRI data share a dependency on the hemodynamic response that intervenes between the observed signals and the actual variations in neuronal population activity (see Section 6.2 for further discussion).

A final distinction is less fundamental to the different approaches and pertains to how they are used in practice. Effective connectivity analysis is often performed as an endpoint of analysis, giving an idea of the interactions between relevant areas over all experimental conditions within a task. However, just as for regional activity levels, it is often in the *modulation* of connectivity by experimental factors that the findings most relevant and interesting can be found. Moreover, because of the dependency on the hemodynamic response, comparing experimental conditions provides a first measure of control for potential regional differences in hemodynamic response characteristics.

#### 4.2. Effective connectivity analyses in multisensory hemodynamic studies

The endeavor to understand how different uni- and multisensory brain areas cooperate to perform integration has commenced this decade. As yet, most studies used PPI (Ethofer et al., 2006; Macaluso et al., 2000; Noesselt et al., 2007; von Kriegstein et al., 2005). Some studies examined interactions between brain areas during unimodal stimulation only (von Kriegstein et al., 2005; Deshpande et al., 2008a), and are therefore not strictly addressing *integration*. A few studies used direction-sensitive effective connectivity methods (Deshpande et al., 2008a; Noesselt et al., 2007; Nop-

peny et al., 2008). In the following, these studies will be described and grouped as to which senses they examined.

##### 4.2.1. Visual-tactile and spatial perception

Using PPI, Macaluso and colleagues (2000) found that the functional coupling between visual cortex (lingual gyrus) and the right inferior parietal lobe and left somatosensory cortex was increased for spatially congruent visuo-tactile stimulation compared to visual-only stimulation. Their interpretation is that tactile input to the somatosensory cortex modulates visual cortex activity through feedback from multisensory areas in the inferior parietal cortex. Sathian and colleagues aimed their connectivity work at understanding the origin of visual cortex activation during tactile perception. In a recent study, they computed directed connectivity simultaneously between 25 regions-of-interest (ROIs) (Deshpande et al., 2008a), using a method based on Granger causality (Deshpande et al., 2008b). The ROIs were selected by haptic shape versus haptic texture perception conditions, and included the visual and somatosensory cortices, prefrontal and posterior parietal cortex. They found evidence for both bottom-up (somatosensory) and top-down (parietal) inputs to visual areas during haptic perception. However, since they used the entire ROI time-courses concatenated over runs and subjects they did not examine modulation of effective connectivity by condition (e.g., shape vs. texture, or haptic vs. visual).

##### 4.2.2. Audiovisual perception

In a recent functional connectivity study, Eckert and colleagues examined functional coupling between auditory and visual cortex during rest (eyes closed) and during visual stimulation (flashing checkerboard) (Eckert et al., 2008). Interestingly, A1 and peripheral V1 were coupled during rest as well as during visual perception. This supports a functional role for connections between auditory cortex and peripheral visual representations, as is also indicated by anatomical connectivity studies (Falchier et al., 2002; Rockland and Ojima, 2003).

Several fMRI studies used PPI to investigate brain circuitry during audiovisual perception of social signals, with somewhat contradicting results. Ethofer and colleagues (2006) studied emotional perception of face-voice pairs, and found no direct coupling between voice and face areas, but only via top-down control from the (supramodal) amygdala. In contrast, Von Kriegstein and colleagues (2005) reported a direct coupling between voice and face areas without involvement of supramodal cortical areas, however, their subjects were engaged in person identification rather than emotional processing. Therefore, involvement of supramodal structures might depend on information type and task goal.

To study how the brain distinguishes temporally synchronous from asynchronous audiovisual information, Noesselt and colleagues (2007) used both PPI analysis and a measure for directed information transfer. Their results revealed that during synchronously presented meaningless patterns, the influence from STS to A1 and V1 was stronger than vice versa, and not significant between A1 and V1 directly. This suggests that STS modulates A1 and V1 when auditory and visual inputs are in synchrony. In a visual-auditory priming study, Noppeney and colleagues (2007) used DCM to investigate the neural mechanism of visuo-auditory incongruity effects for objects and speech. Specifically, they compared two models, one explaining incongruity effects via forward connections (consistent with predictive coding), the other by feedback from cognitive control regions (prefrontal cortex) to lower sensory regions. Bayesian model comparison provided evidence for the feedforward model, supporting the mechanism of predictive coding during speech and object processing.

In sum, the functional and effective connectivity studies described here draw different conclusions about how uni- and

multisensory brain regions cooperate during cross-modal processing or integration, especially with regard to the role of multisensory (also termed *supramodal*) regions. In the next section, we directly examine the role of STS in the integration of letters and speech sounds, by mapping directed interactions to and from STS over the brain.

#### 4.3. Granger causality mapping of letter–speech sound integration

Using the passive event-related data from Van Atteveldt et al. (2007b), effective connectivity was assessed using Granger causality mapping (GCM, Roebroek et al., 2005), which identifies voxels that are sources/targets of directed influences to/from a selected region-of-interest (ROI). Specifically, we tested the hypothesis that the modulation of speech sound processing in the auditory cortex by letters reflects the result of feedback from the STS/STG, as formulated in Sections 3.2 and 3.3. Conform to this hypothesis, we predicted that directed influence from STS to PT is stronger for congruent than for incongruent letter–sound pairs. To test this prediction, we selected ROIs in STS/STG in 10 individual subjects using the conjunction of [auditory > baseline]  $\cap$  [visual > baseline], which identifies heteromodally responsive voxels. This conjunction map was thresholded using the False Discovery Rate (FDR, Genovese et al., 2002) at  $q < 0.05$ . Average Talairach coordinates of the ROI in left STS were  $-54 \pm 1$  (x),  $-40 \pm 2$  (y),  $6 \pm 1$  (z) and the average size was  $213 (\pm 37)$  mm<sup>3</sup>. Using this left STS ROI as reference region (“ref”), single-subject Granger causality maps (GCMs) were calculated for the congruent and incongruent conditions. Next, random-effects group GCMs were created to test modulation of effective connectivity by experimental condition (congruency).

Fig. 2A shows the significant differences between GCMs (GCM modulation map: congruent–incongruent) mapped over the brain with the left STS as reference region (“ref”), showing clusters of modulated connectivity in left and right PT, and right occipital cortex. The GCM modulation map shows significant experimental modulation of directed connectivity with left STS but not the direction of the influence and the condition in which it is stronger. Hence, we plotted the GCM modulation map together with the GCM for congruent and incongruent separately (Fig. 2B). Fig. 2B shows that the GCM modulation map (orange) overlaps with the positive GCM congruent map (dark blue) in auditory cortex marking influence from the left STS to auditory cortex, and not with the GCM incongruent map (no significant voxels). Therefore, the GCM modulation map in Fig. 2A indicates an influence from the left STS to auditory cortex bilaterally and right occipital cortex stronger during congruent letter–sound pairs than incongruent pairs.

In Fig. 2B, the congruency contrast from the activation level analysis (GLM) is also plotted (light blue), which reveals that the auditory cortex regions targeted by STS are the same regions as those in which the congruency effect has been observed in our previous studies using more conventional GLM analyses. Fig. 2C shows the fMRI signal time-courses from the regions revealed with the GCM modulation map, which, interestingly, show the same response pattern as the auditory cortex regions from the earlier fMRI studies (e.g., see Fig. 1). Another interesting result from the GCM analysis is that the region of auditory association cortex targeted by the left STS reference region seems to consist of a posterior and an anterior part that can be distinguished in terms of their response pattern (Fig. 2C): the posterior PT region (PTp) shows a stronger visual response (relative to the auditory) than the more anterior PT region (PTa). In the right PTp, the visual and auditory responses are even equally strong. This unexpected difference demonstrates that the data-driven GCM analysis can reveal additional details that are overlooked by analysis focused on specific contrasts. This is also the case for the occipital region that is tar-

geted by STS and shows a differential response to congruent and incongruent stimulus pairs, but was not revealed using GLM analysis. The location of this occipital target region might correspond to V3/V3A (Tootell et al., 1997).

To conclude, the results from the GCM analyses support our feedback hypothesis: the congruency effect in PT is likely to be caused by differential feedback from STS. Moreover, the GCM results extended the feedback model by revealing that STS might project modulatory feedback to both auditory and visual processing regions. Mapping directed influences to and from the targeted low-level sensory regions is currently in progress to extend the present results. For instance, it may provide additional insights in whether direct visual to auditory influences can also occur, as was suggested by the MMN results (Section 3.6). Furthermore, it might reveal additional sources of influence to the auditory cortex during letter–sound integration.

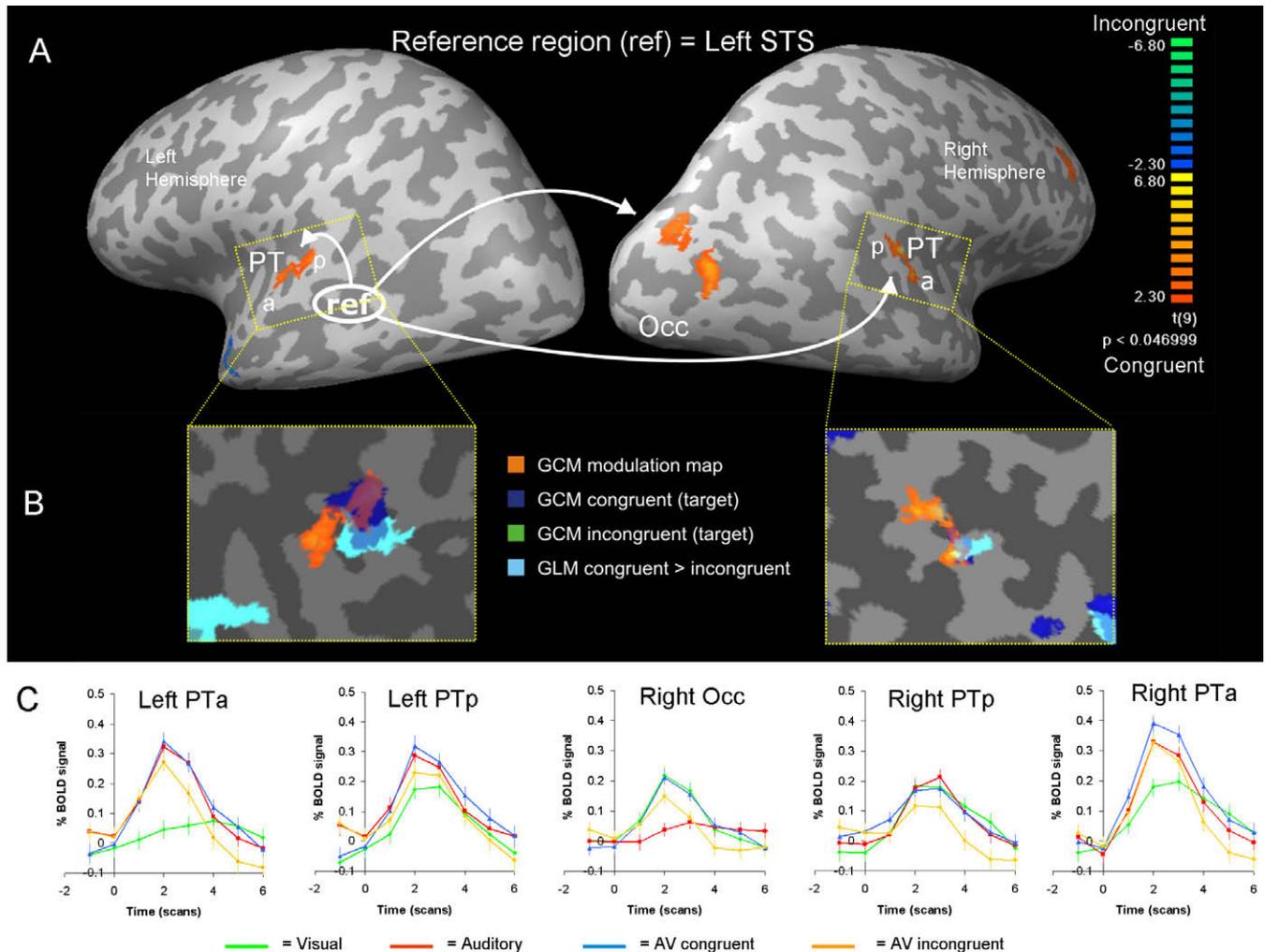
## 5. What have we learned?

### 5.1. Summary and proposed mechanism

The empirical studies reviewed in the previous sections unequivocally demonstrate that the arbitrarily related, but overlearned, pairs of letters and speech sounds can lead to very early visual influences in the auditory cortex, both in processing stage (second level, HS/PT or belt) and in time (around 180 ms after onset speech sound). These influences were expressed as effects of congruency: the response to speech sounds was enhanced when congruent letters were co-presented, and suppressed in the case of incongruent letters. Further notable observations are that these visual influences occurred in absence of a significant response to visual letters in isolation (i.e., not paired with speech sounds), were sensitive to temporal synchrony of the letters and speech sounds, were overruled during an explicit matching task, but not during an auditory task in which integration was implicit. The role of planum temporale in integration of speech and script is also supported by other studies (Buchsbaum et al., 2005; Nakada et al., 2001).

In addition to the auditory cortex, the heteromodal STS/STG was found to play an important role in the integration of letters and speech sounds. It shows convergence of visual and auditory inputs and enhanced responses to bimodal stimulation relative to the unimodal inputs (Fig. 1, right top graph). Furthermore, these enhanced bimodal responses in the heteromodal STS/STG were shown to be more robust over different temporal offsets (Fig. 1, right bottom graph). Although direct influences from visual to auditory processing regions remain an (additional) option, we advocate the interpretation that STS/STG integrates letters and speech sounds, since the unimodal responses converge in this regions whereas the modulated PT/HS regions does not respond to letters alone. Based on the congruency and temporal (a)synchrony, the STS will enhance or suppress speech sound processing in secondary auditory cortex. Our mapping of directed influences using Granger causality provides direct evidence for feedback projections from STS to auditory cortex based on the congruency of the letter–sound pairs (Fig. 2). More indirect support is provided by the finding of stronger directed influence from STS to auditory (and visual) cortex during synchronous versus asynchronous audiovisual patterns (Noesselt et al., 2007). More basic monkey electrophysiology supports the mechanism of visual influences on auditory processing by feedback from superior temporal cortex as well (Ghazanfar et al., 2008; Maier et al., 2008; Schroeder and Foxe, 2002).

In contrast to auditory and heteromodal superior temporal regions, the role of the visual cortex seems to be less robust, but depends more strongly on stimulus quality and sensitivity of the used analysis



**Fig. 2.** Results of the Granger causality mapping analysis. (A) Random-effects ( $n = 10$ ) group GCM difference map (congruent–incongruent). (B) Overlap of GCM modulation map (orange), GCM congruent map (dark blue), and GLM congruency contrast map (light blue) in planum temporale. The GCM incongruent map (green) does not show any significant voxels. (C) Averaged BOLD response time-courses in auditory and visual target regions of STS, selected from the GCM modulation map in A. PTa, anterior planum temporale; PTp, posterior planum temporale; Occ, occipital cortex.

technique. A ventral visual area was found to be involved using degraded inputs (Blau et al., 2008a), whereas a more dorsal visual area was revealed using GCM (Fig. 2). A possible explanation for these different visual areas is given in Section 5.3. Based on the schematic functional neuro-anatomical mechanism proposed in van Atteveldt et al. (2004), we can now extend this schematic to provide a more complete overview of how letter–sound integration might be performed in the literate adult brain, and to gain better insight on which questions remain unanswered (Fig. 3).

### 5.2. A framework to study reading acquisition and dyslexia

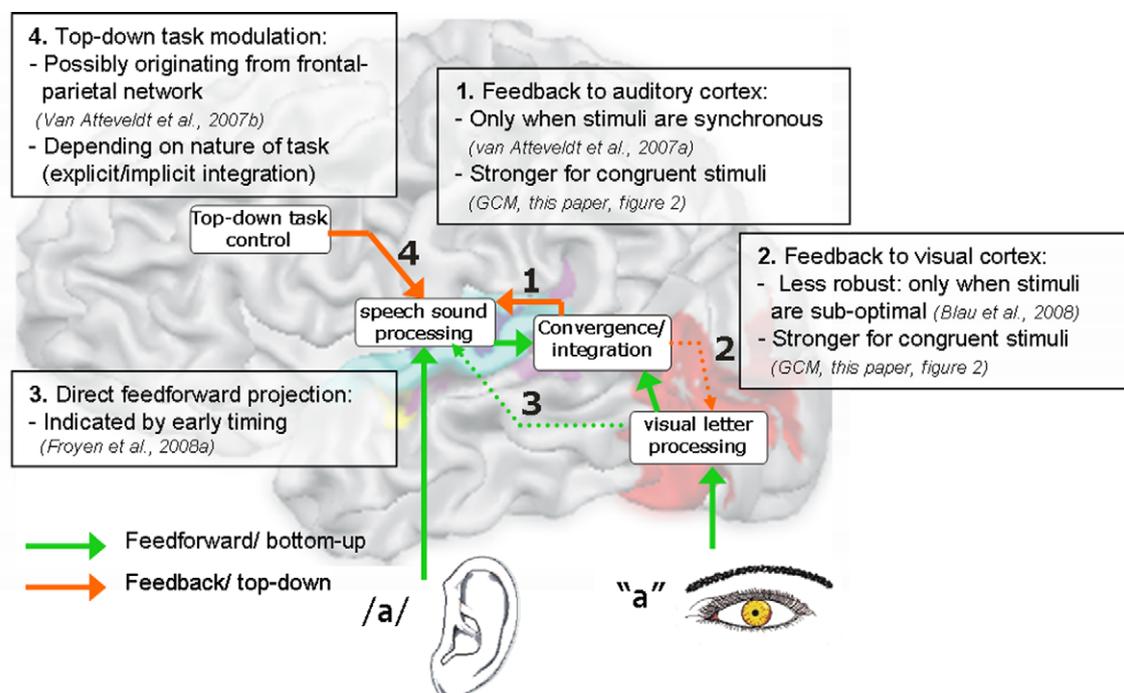
As elucidated in Section 2 and empirically supported by Froyen et al. (2009), literacy learning and automating letter–sound correspondences during acquisition requires a long period of instruction. Interestingly, letter–sound correspondences appear especially hard to learn and automate in developmental dyslexia (Fox, 1994; Snowling, 1980). Although many different causes have been hypothesized, the core deficit in developmental dyslexia is now generally believed to be a phonological processing deficit (Ramus, 2003; Shaywitz and Shaywitz, 2005; Vellutino et al., 2004). A crucial question is how a phonological processing deficit leads to a reading impairment. An interesting, but not yet empirically validated hypothesis is that the phonological deficit impairs

the ability to segment spoken words into phonemes, and therefore to associate letters to the corresponding phonemes (see e.g., Shaywitz and Shaywitz, 2005; Vellutino et al., 2004). In other words, a phonological deficit may result in a reading impairment because letter–sound correspondences are not efficiently learned and/or automated.

The proposed mechanism for normally developed letter–sound integration in healthy adults (Fig. 3) provides a basis to investigate this “missing link” between phonology and reading. Work in progress in our lab indicates that both dyslexic adults (Blau et al., 2009) and children (Blau et al., 2008b) show deviant integration of letters and speech sounds in the auditory cortex and STS/STG. This points to impaired letter–sound integration as fundamental deficit in developmental dyslexia that is present already at the beginning of reading acquisition, and, in spite of years of reading experience, cannot be overcome automatically.

### 5.3. The role of auditory and visual cortices in speech – script integration

The effects of congruency and temporal asynchrony observed in the auditory association cortex were not observed in the visual association cortex (Van Atteveldt et al., 2004, 2007a). Only when



**Fig. 3.** Proposed functional neuro-anatomical mechanism for integration of letters and speech sounds in the “literate” adult brain. Solid lines received converging empirical support; dashed lines are tentative and should be confirmed in future studies. Green lines indicate feedforward projections, orange lines feedback.

using suboptimal visual and auditory stimuli (Blau et al., 2008a), or a different analysis technique (Fig. 2), the visual cortex was also found to be influenced by letter–sound congruency. Following our feedback interpretation, this implies that the outcome of the process of integrating letters and speech sounds is primarily projected back to low-level auditory processing, and less robustly to low-level visual processing areas. Considering the different neural timing profiles within visual and auditory systems, the direction of the modulatory effects between letters and speech sounds may depend on the temporal synchrony of the stimuli. This prediction was however not supported by van Atteveldt et al. (2007a), where we demonstrated that cross-modal modulation of letter processing in the visual association cortex was also absent when letter–sound pairs were presented at different temporal offsets (in both directions).

The observed asymmetrical influence more likely reflects the parasitic nature of written language (see Section 2). Since written language parasites on spoken language and not the other way around, it seems plausible that visual information can access the speech processing system and not vice versa. This also makes sense from a developmental perspective. During reading acquisition, phoneme processing and literacy skills are reciprocally related. Learning the letters of the alphabet and letter–sound mappings strongly influences the processing and representation of single phonemes (Morais and Kolinsky, 2005). This may explain the influence of letters on the processing of isolated speech sounds as observed by the robust congruency effect in the auditory cortex in the reviewed studies.

Interestingly, although the mechanism might not be symmetrical but biased towards auditory processing, this asymmetry seems flexible and might adapt to circumstances such as input quality. In the case of undegraded stimuli, letters are discrete characters with a high signal-to-noise ratio, whereas speech sounds are continuous in time and more variable, and therefore more difficult to recognize (Lieberman, 1992). This preserves the asymmetry because it is more efficient that the unambiguous visual input “helps” auditory processing than vice versa. When visual stimuli are of less

quality, which is the case in Blau et al. (2008a), visual processing also seems to benefit from the combined input.

The analysis of Granger causality during letter–sound integration revealed a different region of visual cortex to be involved: more dorsal occipital cortex, possibly corresponding to area V3/V3a. The two different locations within the visual system indicate that they might be involved in different processes. Campbell described an interesting distinction of processing modes for audiovisual speech with different neural mechanisms (Campbell, 2008): a *complementary* mode, in which visual information complements sub-optimal speech components; and a *correlated* mode, when vision and audition are redundant and provide highly correlated time-varying signals. The putative complementary system was suggested to rely on ventral visual processing areas, the correlated system more on the dorsal stream. Our findings might be explained in light of a similar distinction for the effect of written language on speech processing. When speech is degraded, visual letters provide complementary information, which is consistent with the involvement of the fusiform gyrus (Blau et al., 2008a). The dorsal visual areas revealed by GCM analysis of the data acquired using undegraded speech, hints at the correlated mode. That this activation is less robust (i.e., was not revealed by direct contrasts within the General Linear Model) makes sense because spectro-temporal correlations like in audiovisual speech are much weaker, or even largely absent, in speech–script associations (Amedi et al., 2005; Van Atteveldt et al., 2007a).

## 6. Future directions of multisensory fMRI

While the presented fMRI research has greatly contributed to our knowledge about multisensory processing and integration in the human brain, our knowledge about putative neuronal mechanisms is still rather coarse. For a deeper understanding at the neuronal level, more detailed knowledge about neuronal representations in relevant brain areas would be necessary as well as knowledge on how neurons in the involved areas interact with

each other. There are several potential routes to follow to gain more detailed spatio-temporal information about multisensory processing. Here we discuss mainly the possibilities offered by advanced human fMRI studies, but we will also briefly point out approaches relying on other imaging modalities.

## 6.1. Neuronal coding

### 6.1.1. Increased spatial resolution

Increasing spatial resolution is an obvious approach to obtain more detailed fMRI data, which might help to shed some light on the fine-grained functional organization of small brain areas. High-resolution functional imaging benefits from higher MRI field strength since small (millimeter or even sub-millimeter) voxels still possess a reasonable signal-to-noise ratio. As an example, a 7 T fMRI study showed tonotopic, mirror-symmetric maps within early human auditory cortex (Formisano et al., 2003; see also 3.2). A spatial resolution of 0.5–1 mm might be just enough to resolve cortical columns, which seem to form the basic building blocks (“alphabet”) of complex distributed representations (Fujita et al., 1992; Mountcastle, 1997). Different entities (e.g., letters or vowels) within a specific area could be coded by a specific distributed activity pattern across cortical columns. If this reasoning is correct, an important research strategy would aim to unravel the specific alphabet in various parts of the cortex, including auditory cortex and heteromodal areas in the STS/STG. Recently it became indeed possible to reliably measure orientation columns in the human primary visual cortex using high-field (7 T) fMRI (Yacoub et al., 2008).

### 6.1.2. Multi-voxel pattern analysis (MVPA)

While standard analyses using the General Linear Model (GLM) process the time-course of each voxel independently, and data-driven methods, such as Independent Component Analysis (ICA) search for functional networks in the whole four-dimensional data set, several new analysis approaches focus on discriminating more local, distributed patterns using multi-voxel pattern analysis (MVPA) techniques (e.g., De Martino et al., 2008; Haxby et al., 2001; Haynes and Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte et al., 2006). In these approaches, data from voxels within a region are jointly analyzed; distributed activity patterns corresponding to different conditions are statistically compared using multivariate approaches (e.g., MANOVA) or machine learning tools (e.g., Support Vector Machines, SVMs). SVMs are trained to classify evoked responses from a subset of the data while leaving some data (or entirely new data) aside for testing generalization performance of the trained classifier. By properly weighting the contribution of individual voxels in a region, (local) multivariate pattern analysis opens the possibility to separate spatially distributed patterns, which would be potentially too weak to be discovered by single-voxel analysis. Recently, it has been shown (Formisano et al., 2008) that distributed patterns extending from early auditory cortex into STS/STG contain enough information to reliably separate distributed responses evoked by individual vowels spoken by different speakers from each other; new exemplars of vowels were classified correctly even if they were spoken by a novel speaker. Importantly, this discriminative power could not be observed when analyzing responses from single-voxels. In the context of multisensory processing it would be interesting to teach classifiers to discriminate responses from bimodal letter–sound stimuli in order to obtain more information about how specific stimulus combinations are represented in auditory cortex (e.g., PT) and STS/STG; different training signals (classification labels) could be used for the same fMRI data to either train to discriminate letter–sound pairs or to extract only the letter or sound of a pair. These different training tasks should reveal which parts in the cor-

tical network would more prominently code for letters, sounds or the combination of both; Furthermore, if learning and generalization would be successful, the identified representations would allow to predict from a single trial response which specific letter–sound combination was presented to the subject. Such knowledge would be highly relevant to build computational models of multisensory processing (see Section 6.3).

### 6.1.3. The power of fMRI designs

Besides pushing the limits of fMRI spatial and temporal resolution and using sensitive multivariate analysis tools, the invention of ingenious fMRI designs might also help to reveal details of neuronal coding at more standard field strengths (e.g., 3 T). The fMRI adaptation paradigm (Grill-Spector and Malach, 2001), for example, may be used to gain insights into sub-voxel neuronal representations even when standard-resolution imaging is used. fMRI adaption (fMR-A) exploits the repetition suppression effect, i.e., the observation that the response to repeated presentations of a stimulus evokes a weaker response than the first presentation. In the fMR-A paradigm, responses to sequences of two or more stimuli are compared, which are identical in the control condition, but vary in certain stimulus properties in the main condition(s). The obtained response amplitudes in the different conditions can be used to assess to what extent the same neurons are responding to the “repeated” stimulus presentation (suppressed response) within a voxel. fMR-A could be used to distinguish uni- and multisensory neuronal populations within voxels, and to further investigate the functional properties of multisensory subpopulations (van Atteveldt et al., 2008).

Furthermore, the results of the SOA experiment (Section 3.3) demonstrate that effects of time can be revealed using fMRI despite the sluggishness of the measured hemodynamic response. For example, fMRI response strength to a whole range of SOAs can be used to verify the phase resetting hypothesis from monkey electrophysiology (Lakatos et al., 2007). However, it should be noted that pure phase resetting results in pre- to poststimulus phase synchronization without an accompanying power increase, indicating that whereas patterns of better and worst SOAs corresponding to certain oscillatory frequencies can be confirmed in humans using fMRI, the underlying phase resetting mechanism can not directly be verified.

## 6.2. Dynamic processes and neuronal interactions

Rapid changes in neuronal activity cannot be detected with fMRI since successive temporal events lead to an integrated BOLD response. This poses a severe obstacle for the interpretation of a major finding of our studies: Is the congruency effect observed within PT calculated directly there (i.e., bottom-up) or is the observed effect the result of top-down influences from higher hetero-modal areas? As has been discussed above (Section 4), Granger causality mapping (GCM) is an interesting approach to solve such questions using fMRI data. By exploiting temporal precedence, this effective connectivity tool has the potential to estimate the direction of influences between brain areas directly from voxel’s time-course data. The performed GCM analysis favored indeed the interpretation that the congruency effect in auditory cortex results from top-down stimulation from STS/STG cortex. While encouraging, the results of the fMRI-based GCM is not conclusive as it might still miss the dynamic interactions beyond the fMRI temporal resolution (see Sections 3.2 and 4.3). Therefore, it would be beneficial to support this interpretation with additional data from other imaging modalities with a high temporal resolution such as EEG or MEG. Unfortunately, it remains a difficult problem to reliably separate closely spaced electrical sources from each other, e.g., PT and STS/STG, despite progress in fMRI con-

straint source modeling (e.g., Goebel and Esposito, 2009; Lin et al., 2006). Direct intra-cranial electrical recordings from the human brain might, thus, be necessary to finally resolve the issue whether the congruency effect in early auditory cortex is the result of intrinsic processing or whether it is the result of top-down signals from heteromodal cortex in the STS/STG.

### 6.3. Towards computational modeling of multisensory processing

The presented data from a series of fMRI experiments as well as the effective connectivity results emerging from Granger causality mapping have been interpreted in a summary model (5.1, Fig. 3) identifying the likely role of brain areas involved in letter–speech sound integration as well as the information flow between these areas. As has been highlighted in the previous sections, data might soon be available providing further constraints at the representational level of individual letters, sounds and letter–sound combinations. In light of the richness of the present and potential future data, it seems to become feasible to build computational models to further specify hypotheses of neuronal coding and neuronal interactions. We have selected the framework of large-scale recurrent neural networks (Goebel and De Weerd, 2009) to more precisely model our ideas in the future. We have selected neural network models because they allow to clearly specifying assumptions about neuronal representations as well as structural connectivity within and between simulated brain areas. Since such models are formulated at a (simplified) neuronal level, they are also able to incorporate results from electrophysiological animal studies. Furthermore, these network models allow predicting the spatio-temporal implications of the implemented representational and structural assumptions by feeding the networks with relevant unimodal and bimodal auditory and visual stimuli. Running simulations of implemented models may help to understand more precisely how the congruency effect in PT “emerges” from synaptic influences from modeled multisensory neurons in cortical columns of interconnected heteromodal brain areas.

### Acknowledgements

This work is supported by the Dutch Organization for Scientific Research (NWO, VENI Grant # 451-07-020 to N.v.A.) and the European Community's Seventh Framework Programme ([FP/2007-2013] Grant # 221187 to N.v.A.).

### References

- Amedi, A., von Kriegstein, K., Van Atteveldt, N.M., Beauchamp, M.S., Naumer, M.J., 2005. Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research* 166, 559–571.
- Beauchamp, M., 2005. Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3, 93–113.
- Blau, V., Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2008a. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *European Journal of Neuroscience* 28, 500–509.
- Blau, V., van Atteveldt, N.M., Seitz, J., Goebel, R., Blomert, L., 2008b. Bridging the gap between speech perception and reading: evidence from pediatric neuroimaging. *Journal of Cognitive Neuroscience* (Suppl.), S140.
- Blau, V., Van Atteveldt, N., Ekkebus, M., Goebel, R., Blomert, L., 2009. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology* 19, 503–508.
- Buchel, C., Friston, K., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* 7, 768–778.
- Buchel, C., Friston, K., 2000. Assessing interactions among neuronal systems using functional neuroimaging. *Neural Networks* 13, 871–882.
- Buchsbaum, B.R., Olsen, R.K., Koch, P.F., Kohn, P., Shane Kippenhan, J., Faith Berman, K., 2005. Reading, hearing, and the planum temporale. *NeuroImage* 24, 444–454.
- Budinger, E., Heil, P., Hess, A., Scheich, H., 2006. Multisensory processing via early cortical stages: connections of the primary auditory cortical field with other sensory systems. *Neuroscience* 143, 1065–1083.
- Byrne, B., Fielding-Barnsley, R., Ashley, L., 1996. What does a child bring to the task of learning to read? A summary of the New England Reading Acquisition Projects. *Australian Journal of Psychology* 48, 119–123.
- Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex* 11, 1110–1123.
- Calvert, G.A., Brammer, M.J., Bullmore, E.T., Campbell, R., Iversen, S.D., David, A.S., 1999. Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10, 2619–2623.
- Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. *Trends in Cognitive Sciences* 11, 535–543.
- Campbell, R., 2008. The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 363, 1001–1010.
- Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. *NeuroImage* 22, 466–476.
- Council, N.R., 1998. Preventing Reading Difficulties in Young Children. National Academy Press, Washington, DC.
- De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E., 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage* 43, 44–58.
- Deshpande, G., Hu, X., Stilla, R., Sathian, K., 2008a. Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *NeuroImage* 40, 1807–1814.
- Deshpande, G., Laconte, S., James, G.A., Peltier, S., Hu, X., 2008b. Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping* Jun 6. [Epub ahead of print].
- Driver, J., Noesselt, T., 2008. Multisensory interplay reveals crossmodal influences on ‘sensory-specific’ brain regions, neural responses, and judgments. *Neuron* 57, 11–23.
- Eckert, M.A., Kamdar, N.V., Chang, C.E., Beckmann, C.F., Greicius, M.D., Menon, V., 2008. A cross-modal system linking primary auditory and visual cortices: evidence from intrinsic fMRI connectivity analysis. *Human Brain Mapping* 29, 848–857.
- Eden, G.F., Moats, L., 2002. The role of neuroscience in the remediation of students with dyslexia. *Nature Neuroscience* 5, 1080–1084.
- 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. 135–154.
- Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., Reiterer, S., Grodd, W., Wildgruber, D., 2006. Impact of voice on emotional judgment of faces: an event-related fMRI study. *Human Brain Mapping* 27, 707–714.
- Falchier, A., Clavagnier, S., Barone, P., Kennedy, H., 2002. Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience* 22, 5749–5759.
- 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.
- Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science* 322, 970–973.
- Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., Goebel, R., 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40, 859–869.
- Fox, E., 1994. Grapheme–phoneme correspondence in dyslexic and matched control readers. *British Journal of Psychology* 85, 41–53.
- Friston, K., 1994. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping* 2, 56–78.
- Friston, K., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *NeuroImage* 19, 1273–1302.
- Friston, K.J., Frith, C.D., Frackowiak, R.S., 1993a. Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping* 1, 69–79.
- Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S., 1993b. Functional connectivity: the principle component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism* 13, 5–14.
- Friston, K.J., Buchel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6, 218–229.
- 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. *Neuroscience Letters* 430, 23–28.
- Froyen, D.J., Bonte, M.L., Van Atteveldt, N., Blomert, L., 2009. The long road to automation: neurocognitive development of letter–speech sound processing. *Journal of Cognitive Neuroscience* 21, 567–580.
- Fujita, I., Tanaka, K., Ito, M., Cheng, K., 1992. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360, 343–346.
- Genovese, C., Lazar, N., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878.
- Ghazanfar, A.A., Chandrasekaran, C.F., 2007. Paving the way forward: integrating the senses through phase-resetting of cortical oscillations. *Neuron* 53, 162–164.
- Ghazanfar, A.A., Chandrasekaran, C., Logothetis, N.K., 2008. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *Journal of Neuroscience* 28, 4457–4469.
- Gleitman, L.R., Rozin, P., 1977. The structure and acquisition of reading I: relations between orthographies and the structure of language. In: Reber, A., Scarborough, D. (Eds.), *Towards a Psychology of Reading: The Proceedings of the CUNY Conferences*. Lawrence Erlbaum Associates, Hillsdale, NJ.

- Goebel, R., Esposito, F., 2009. The added value of EEG-fMRI in imaging neuroscience. In: Mulert, C., Lemieux, L. (Eds.), *Combined fMRI-EEG Data Analysis*. Springer, Berlin.
- Goebel, R., De Weerd, P., 2009. Perceptual filling-in: from experimental data to neural network modeling. In: Gazzaniga, M.S. (Ed.), *The Cognitive Neurosciences IV*. MIT Press, Cambridge.
- Gough, P.B., Hillinger, M.L., 1980. Learning to read: an unnatural act. *Bulletin of the Orton Society: An Interdisciplinary Journal of Specific Language Disability* 30, 179–196.
- Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 100, 253–258.
- Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica* 107, 293–321.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.
- Haynes, J.D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience* 8, 686–691.
- Herdman, A.T., Fujioka, T., Chau, W., Ross, B., Pantev, C., Picton, T.W., 2006. Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neuroscience Letters* 399, 61–66.
- Kaas, J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America* 97, 11793–11799.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 8, 679–685.
- Kayser, C., Logothetis, N.K., 2007. Do early sensory cortices integrate cross-modal information? *Brain Structure and Function* 212, 121–132.
- Kayser, C., Petkov, C.I., Augath, M., Logothetis, N.K., 2007. Functional imaging reveals visual modulation of specific fields in auditory cortex. *Journal of Neuroscience* 27, 1824–1835.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America* 103, 3863–3868.
- Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., Schroeder, C.E., 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292.
- Laurienti, P.J., Perrault, T.J., Stanford, T.R., Wallace, M.T., Stein, B.E., 2005. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research* 166, 289–297.
- Lieberman, A.M., 1992. The relation of speech to reading and writing. In: Frost, R., Katz, L. (Eds.), *Orthography, Phonology, Morphology and Meaning*. Elsevier Science Publishers B.V., Amsterdam, pp. 167–178.
- Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* 31, 160–171.
- Macaluso, E., Frith, C.D., Driver, J., 2000. Modulation of human visual cortex by crossmodal spatial attention. *Science* 289, 1206–1208.
- Maier, J.X., Chandrasekaran, C., Ghazanfar, A.A., 2008. Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology* 18, 963–968.
- Mattingly, I.G., 1972. Reading, the linguistic process, and linguistic awareness. In: Kavanagh, J.F., Mattingly, I.G. (Eds.), *Language by Ear and by Eye: The Relationship between Speech and Reading*. MIT Press, Cambridge, MA.
- McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. *Human Brain Mapping* 2, 2–22.
- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping* 6, 160–188.
- Molholm, S., Foxe, J.J., 2005. Look 'hear', primary auditory cortex is active during lip-reading. *Neuroreport* 16, 123–124.
- Morais, J., Kolinsky, R., 2005. Literacy and cognitive change. In: Snowling, M., Hulme, C. (Eds.), *The Science of Reading – A Handbook*. Blackwell Publishing, Oxford.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* 13, 684–701.
- Mountcastle, V.B., 1997. The columnar organization of the neocortex. *Brain* 120, 701–722.
- Nakada, T., Fujii, Y., Yoneoka, Y., Kwee, I.L., 2001. Planum temporale: where spoken and written language meet. *European Neurology* 46, 121–125.
- Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.-J., Driver, J., 2007. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience* 27, 11431–11441.
- Noppeney, U., Josephs, O., Hocking, J., Price, C.J., Friston, K.J., 2008. The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex* 18, 598–609.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I.P., Mottonen, R., Tarkiainen, A., Sams, M., 2005. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16, 125–128.
- Perfetti, C.A., Sandak, R., 2000. Reading optimally builds on spoken language: implications for deaf readers. *Journal of Deaf Studies and Deaf Education* 5, 32–50.
- Perrault, T.J.J., Vaughan, J.W., Stein, B.E., Wallace, M.T., 2005. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *Journal of Neurophysiology* 93, 2575–2586.
- Petkov, C.I., Kayser, C., Augath, M., Logothetis, N.K., 2006. Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biology* 4, e215.
- Picton, T.W., Alain, C., Otten, L., Ritter, W., Achim, A., 2000. Mismatch negativity: different water in the same river. *Audiology and Neuro-otology* 5, 111–139.
- Polk, T., Stallcup, M., Aguirre, G., Alsop, D., D'Esposito, M., Detre, J., Farrah, M., 2002. Neural specialization for letter recognition. *Journal of Cognitive Neuroscience* 14, 145–159.
- Price, C.J., Mechelli, A., 2005. Reading and reading disturbance. *Current Opinion in Neurobiology* 15, 231–238.
- Raij, T., Uutela, K., Hari, R., 2000. Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625.
- Ramus, F., 2003. Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology* 13, 212–218.
- Riecke, L., van Opstal, A.J., Goebel, R., Formisano, E., 2007. Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *Journal of Neuroscience* 27, 12684–12689.
- Rockland, K.S., Ojima, H., 2003. Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology* 50, 19–26.
- Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. *NeuroImage* 25, 230–242.
- Schroeder, C.E., Foxe, J.J., 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research* 14, 187–198.
- Schroeder, C.E., Foxe, J.J., 2005. Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology* 15, 1–5.
- Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences* 2, 106–113.
- Schröger, E., 1998. Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers* 30, 131–145.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Shaywitz, S.E., Shaywitz, B.A., 2005. Dyslexia (specific reading disability). *Biological Psychiatry* 57, 1301–1309.
- Snowling, M., Hulme, C. (Eds.), 2005. *The Science of Reading: A Handbook*. Blackwell Publishing, Oxford.
- Snowling, M.J., 1980. The development of grapheme-phoneme correspondence in normal and dyslexic readers. *Journal of Experimental Child Psychology* 29, 294–305.
- Stein, B.E., Meredith, M.A., 1993. *The Merging of the Senses*. MIT Press, Cambridge, MA.
- Tootell, R.B., Mendola, J., Hadjikhani, N.K., Ledden, P.J., Liu, A.K., Reppas, J.B., Sereno, M.I., Dale, A.M., 1997. Functional analysis of V3A and related areas in human visual cortex. *Journal of Neuroscience* 17, 7060–7078.
- 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., Blau, V., Blomert, L., Goebel, R., 2008. fMR-adaptation reveals multisensory integration in human superior temporal cortex. In: *Annual Meeting of the International Multisensory Research Forum*, Hamburg, Germany.
- Van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2007a. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex* 17, 962–974.
- Van Atteveldt, N.M., Formisano, E., Goebel, R., Blomert, L., 2007b. Top-down task effects override automatic multisensory responses to letter-sound pairs in auditory association cortex. *NeuroImage* 36, 1345–1360.
- van de Ven, V.G., Formisano, E., Prvulovic, D., Roeder, C.H., Linden, D.E., 2004. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Human Brain Mapping* 22, 165–178.
- 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? *Journal of Child Psychology and Psychiatry* 45, 2–40.
- von Kriegstein, K., Kleinschmidt, A., Sterzer, P., Giraud, A., 2005. Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience* 17, 367–376.
- Yacoub, E., Harel, N., Ugurbil, K., 2008. High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences of the United States of America* 105, 10607–10612.