

Multisensory functional magnetic resonance imaging: a future perspective

Rainer Goebel · Nienke van Atteveldt

Received: 24 March 2009 / Accepted: 25 May 2009 / Published online: 17 June 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Advances in functional magnetic resonance imaging (fMRI) technology and analytic tools provide a powerful approach to unravel how the human brain combines the different sensory systems. In this perspective, we outline promising future directions of fMRI to make optimal use of its strengths in multisensory research, and to meet its weaker sides by combining it with other imaging modalities and computational modeling.

Introduction

The potential to measure whole-brain volumes in about 2 s and its non-invasiveness make functional magnetic resonance imaging (fMRI) indispensable in human cognitive neuroscientific research. fMRI enables mapping large-scale brain activation (Huettel et al. 2004) as well as interaction patterns (Friston et al. 1997; Roebroeck et al. 2005), which

yield essential exploratory knowledge on brain functioning at the systems level (Logothetis 2008). Interestingly, these mappings are not restricted to the cortex, but may for instance include cortical–subcortical interaction patterns. This is highly relevant since in addition to the superior colliculi (Stein and Meredith 1993), the thalamus and thalamo-cortical interactions in particular may be important for multisensory integration (Schroeder et al. 2003; Hackett et al. 2007; Cappe et al. 2009).

A weak point is that the hemodynamic nature of the fMRI signal (the “Blood Oxygenation Level Dependent” or BOLD signal) makes it an indirect and relatively sluggish measure, and therefore inadequate to capture fast dynamic neural processes. Methods measuring human brain activity more directly and with temporal resolution in the range of neural dynamics, such as scalp and intracranial electroencephalography (EEG) and magneto-encephalography (MEG), provide only limited coverage and have lower spatial accuracy (except intracranial EEG, but this method is invasive and depends on patient populations). Therefore, the advantages of fMRI should be optimally exploited and combined with these complementary methods. In the present perspective, we will outline recent advancements in fMRI technology, design and analytical approaches that can promote a deeper understanding of our brain’s ability to combine different sensory systems.

First, we will discuss why multisensory research poses extra challenges on how to interpret fMRI results at the neuronal level. A major challenge is the problem of choosing appropriate (statistical) criteria for deciding to what extent a voxel or region is involved in integration. Typically, fMRI studies on multisensory integration compare fMRI responses to multisensory stimulation (e.g., audiovisual) to their unisensory counterparts (separate auditory and visual stimuli), using univariate General

R. Goebel (✉) · N. van Atteveldt
Department of Cognitive Neuroscience,
Faculty of Psychology and Neuroscience, Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands
e-mail: R.Goebel@maastrichtuniversity.nl

N. van Atteveldt
e-mail: N.vanAtteveldt@maastrichtuniversity.nl

R. Goebel
Netherlands Institute for Neuroscience,
Meibergdreef 47, 1105 BA Amsterdam, The Netherlands

N. van Atteveldt
Department of Psychiatry,
Columbia University College of Physicians and Surgeons
and New York State Psychiatric Institute, 1051 Riverside Drive,
10032 New York, NY, USA

Linear Models (GLMs) at each voxel (Friston et al. 1994). Because none of the proposed metrics of multisensory integration (see below) that can be applied to estimated beta values is ideal, alternative designs and analytical tools need to be explored. One alternative type of design is to make use of repetition suppression effects, as is done in fMRI-adaptation (Grill-Spector and Malach 2001). Other designs that can offer more flexibility are multifactorial designs in which multiple factors can be simultaneously manipulated, e.g., semantic and temporal correspondence of multisensory inputs (Van Atteveldt et al. 2007), or both within- and between-group factors (Blau et al. 2009). While these approaches are based on voxel-wise estimates, multi-voxel pattern analysis (MVPA) approaches (Haynes and Rees 2005; De Martino et al. 2008) jointly analyze data from multiple voxels within a region. By focusing on distributed activity patterns, this approach opens the possibility to separate and localize spatially distributed patterns, which would be potentially too weak to be detected by single-voxel (univariate) analysis. As recently applied to classify sensory-motor representations (Etzel et al. 2008), it will be very interesting to apply MVPA analogously to sensory-sensory representations to test whether representations of events in one modality generalize to other modalities.

Besides methodological improvements, the potential benefits of technological advancements will be discussed. Scanners with ultra-high magnetic field strengths (≥ 7 Tesla) provide enough signal-to-noise for functional scanning at sub-millimeter spatial resolution, which may allow to directly map distributed representations at the columnar level (Yacoub et al. 2008). To exploit higher spatial resolution scans also at the group-level, we will discuss the use of advanced, cortex-based, multi-subject alignment tools, which match corresponding macro-anatomical structures (gyri and sulci) across subjects. Finally, since the eventual goal is to understand dynamic processes and neuronal interactions, which take place in a millisecond time-scale, advancements in combining fMRI with more “temporal” methods will be outlined. For example, because of its whole-brain coverage and non-invasive nature, fMRI can be used to raise specific new predictions that can be verified by other methods such as human intracranial recordings which have both spatial and temporal resolution, but limited coverage and practical restraints. A relevant example is a recent intracranial study demonstrating the cortical dynamics of audiovisual speech processing (Besle et al. 2008), testing predictions raised by previous fMRI (lacking temporal precision) and scalp EEG studies (lacking spatial precision). We will conclude by discussing the role of computational modeling in integrating results from multisensory neuroimaging experiments in a common framework.

Statistical inference in multisensory fMRI: how to define integration?

Deciding whether a neuron is “multisensory” on basis of single cell recordings is relatively straightforward, using directly acquired data on how the recorded neuron responds to different types of stimulation (unisensory, multisensory). Integration is thought to occur when the response to a combined stimulus (e.g., audiovisual) is different from the response predicted on basis of the separate responses (e.g., auditory and visual). The initially employed criterion is that a neuron’s spike count during multisensory stimulation should exceed that to the most effective unisensory stimulus (Stein and Meredith 1993). An interesting observation is that some multisensory neurons respond super-additively: the response to multisensory stimuli not only exceeds the maximal unisensory response, but even the summed (or additive) response to both (or multiple) sensory modalities (Wallace et al. 1996).

When dealing with fMRI data, the decision of when a voxel or region is multisensory is far more complicated. An important reason is that instead of single neurons (or small units), the responses of several hundred thousand neurons are combined in the signal of one fMRI unit (voxel). This is a problem because voxels are quite unlikely to consist of homogeneous neural populations. Instead, the large sample of neurons can be made up of mixed unisensory and multisensory sub-populations (Laurienti et al. 2005), and multisensory sub-populations on their turn can consist of multisensory neurons with very diverse response properties (additive, super- or sub-additive; Perrault et al. 2005). Therefore, the voxel-level response can have many different origins at the neural level. For example, an enhanced BOLD response for multisensory relative to unisensory stimulation can be due to “true” multisensory neurons integrating stimulation from two or more sensory modalities, but it can just as well be explained by driving two unisensory sub-populations instead of one. If the latter scenario would be true, one might wrongly infer multisensory integration at the neuronal level. A super-additive BOLD response is less prone to such false inferences (Calvert 2001), but it is unlikely to be observed because of the same heterogeneity of response types (unisensory, super-additive, sub-additive) that may cancel each other out at the voxel level (Beauchamp 2005b; Laurienti et al. 2005). Observation of an enhanced (whether super-additive or not) BOLD response during multisensory stimulation, therefore, has to be carefully interpreted and will most likely be based on a mixture of multisensory and unisensory responding neurons.

Moreover, the BOLD response does not increase linearly with increasing neuronal population activity but reaches a ceiling level, i.e., it saturates (Buxton et al. 2004; Haller et al. 2006). Whereas the dynamic range of single neurons

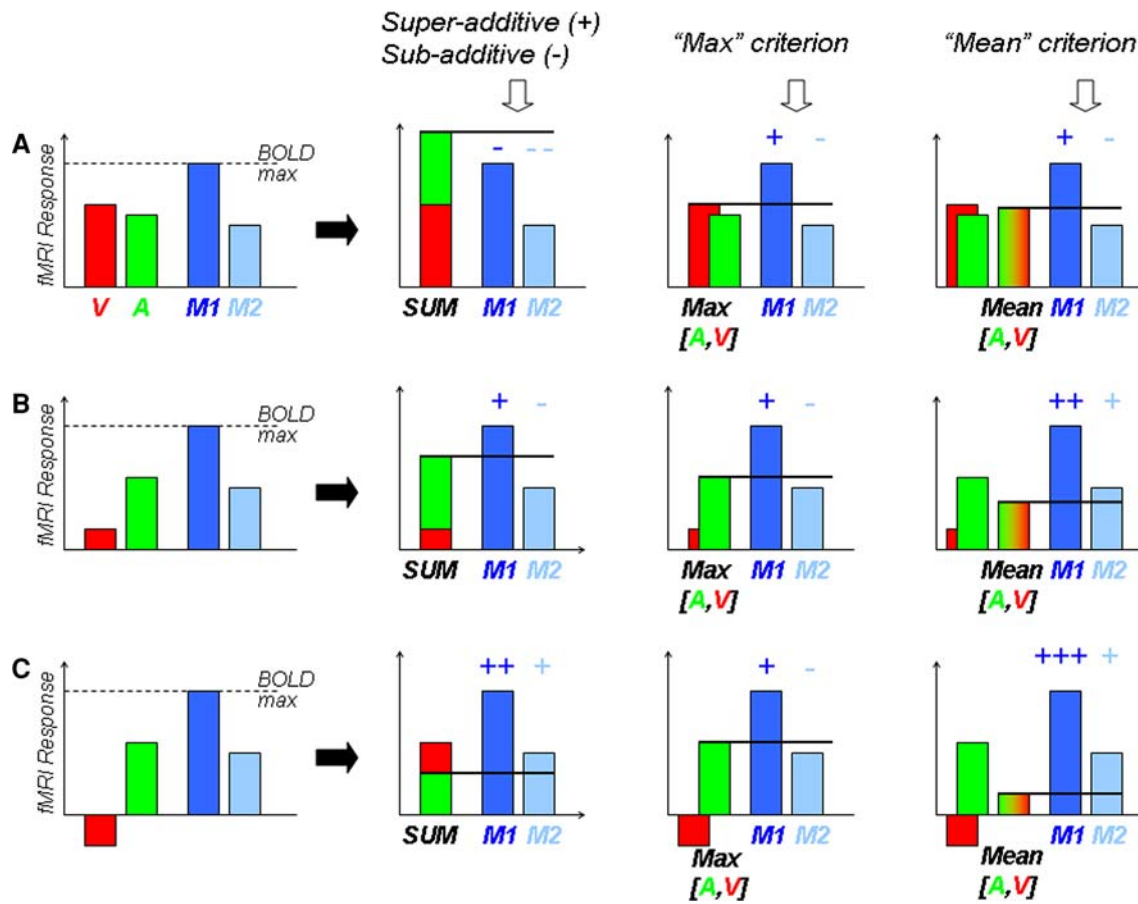


Fig. 1 Classification by different statistical criteria (columns) for hypothetical brain regions with different unisensory (fMRI) response profiles (a–c). **a** Heteromodal response: a significant response to both unisensory stimulation modalities (auditory and visual). **b** Auditory-specific response and a weak visual response. **c** Auditory-specific response and a negative visual response. Bars indicate the fMRI activation level for different unisensory and multisensory stimulation conditions: visual (V red), auditory (A green), and two different audio-visual/multisensory conditions (M1 dark blue; M2 light blue). The dotted

line in the first column (“BOLD max”) represents the maximal fMRI response due to hemodynamic saturation. The solid lines in columns 2–4 represents the classification criterion: summed unisensory activation level (A + V) for the super-additivity criterion, maximal unisensory activation level ([A, V]max) for the “Max” criterion, and mean unisensory activation level (A + V)/2 for the “Mean” criterion. Plus and minus symbols indicate classification type (super-additivity/enhancement vs. sub-additivity/suppression) and strength

can have intrinsic functional properties (Perrault et al. 2005), the limited dynamic range of the BOLD response is a characteristic of the vascular system (for instance, limited capability of vessel dilation) and therefore confounds neurofunctional interpretations. In other words, BOLD saturation might conceal increased neuronal population responses to multisensory stimulation, especially when unisensory stimuli already evoke substantial responses (Fig. 1a). This may result in false negatives, since integration at the neuronal level is not well reflected at the voxel level.

Statistical criteria: different classifications in different situations

When using fMRI studies to identify multisensory integration regions in the human brain, we have to seek for means

to objectively define integrative fMRI responses. Several statistical criteria have been suggested to infer multisensory integration from fMRI data (Calvert 2001; Beauchamp 2005b; Laurienti et al. 2005; Driver and Noesselt 2008; Stevenson et al. 2008) ranging from stringent to liberal, respectively: the criterion of super-additivity, the max criterion and the mean criterion. The super-additivity criterion states that the multisensory response should exceed the sum of the unisensory responses to be defined as integrative. The max criterion is defined in analogy to the criterion used to infer multisensory enhancement or suppression on the single neuron level (Stein and Meredith 1993) and states that the multisensory fMRI response should be stronger than the most effective unimodal response. The most liberal criterion is the mean criterion, stating that the multisensory response should exceed the mean of the unimodal responses. Typically, integration is defined by a positive

outcome using any of the criteria (super-additivity or enhancement); in this case the stimuli are assumed to “belong together”. A negative outcome is typically interpreted as inhibited processing (sub-additivity or suppression), which can be viewed as another type/direction of integration, for stimuli that are assumed to “not belong together”. No difference between multi- and unisensory responses (additivity, no interaction) is interpreted as no integration, in case two inputs do not influence each other’s processing in that voxel or region.

To gain insight in how the different criteria reach their classifications, Fig. 1 illustrates their outcomes with respect to specific multisensory (“M”) responses, in regions with different unisensory (visual and auditory) response profiles. Region “A” shows a heteromodal response, i.e., the area responds significantly to both unisensory stimulation types. This profile is, for instance, typical for regions in the posterior superior temporal sulcus (STS; see Amedi et al. 2005; Beauchamp 2005a). Regions “B” and “C” show a sensory-specific (auditory) response (typical for areas in auditory cortex), with a weak visual response in “B” and a negative visual response in “C”. The three integration criteria are applied to the fMRI activity for two different multisensory (audiovisual) stimulus types “M1” and “M2”. M1 evokes a strong fMRI response, higher than each of the unisensory responses, whereas M2 evokes a much weaker response that does not exceed either of the unisensory responses. As will be discussed below, the figure shows that unisensory response profiles as well as the BOLD response saturation level (“BOLD max”) both affect classification of the fMRI response to M1 and M2 differently for the three criteria.

In region “A”, the sum of the two unisensory responses exceeds the BOLD saturation level implying that no multisensory response can show super-additivity. As a consequence, both M1 and M2 responses are classified as sub-additive (–), and hence as not or negatively “integrative”, even though M1 is clearly boosted and M2 is not. Both the max- and the mean criteria classify the response to M1 as enhanced (+) and M2 as suppressed (–) relative to unisensory responses. In region “B”, the summed response does not exceed the BOLD saturation level and hence the response to M1 is now classified as super-additive (+), and the weak M2 response as sub-additive (–). The max criterion again classifies the M1 response as enhanced (+) and the M2 response as suppressed (–). In contrast, the mean criterion classifies both M1 and M2 as enhanced (+), while the M2 response does not exceed the auditory response. In region “C”, the summed response is actually lower than each of the unisensory responses because one of the responses is negative. The super-additivity criterion classifies both M1 and M2 responses as super-additive (+), but it is questionable how meaningful it is to sum the responses when one is negative (Calvert 2001). The mean criterion also classifies both M1

and M2 as enhanced, whereas the max criterion still classifies M1 as enhanced (+) and M2 as suppressed (–).

The super-additivity criterion seems to be prone to false negatives in regions such as “A” due to BOLD saturation, and possibly to false positives in “C” due to a negative response in one of the modalities. The first bias can be limited by using “weak” stimuli to prevent BOLD saturation (Calvert 2001; Stevenson et al. 2008). Note that stimuli at detection threshold are also recommended from a neural perspective (inverse effectiveness; Stein and Meredith 1993), since such stimuli increase the need for integration. On the other extreme, the mean criterion seems to be too liberal especially when one of the unisensory responses is weak (“B”) or negative (“C”), which reduces the mean in such a way that a multisensory response exceeds the mean even when weaker than the largest unisensory response. Therefore, the mean criterion can be misleading, especially when examining low-level sensory regions such as the auditory cortex.

In sum, whereas saturation confounds can be avoided by presenting weak stimuli, both super-additivity and mean criteria seem biased toward classifying a multisensory response as integrative in sensory-specific brain regions (like “B” or “C”). This is problematic because many recent studies support involvement of low-level sensory-specific brain regions in multisensory integration (reviewed in Schroeder and Foxe 2005; Ghazanfar and Schroeder 2006; Macaluso 2006; Kayser and Logothetis 2007; Driver and Noesselt 2008). As argued in the introduction, an asset of fMRI is that functional maps can be created over the whole brain. In such whole-brain analyses, identical statistical tests are performed in all sampled voxels. Therefore, it is important that a criterion for multisensory integration is suitable in all voxels, regardless of different unisensory response profiles. The classification based on the max criterion seems most robust to different unisensory response profiles. It can also be argued that this is a disadvantage, because the classification by itself does not give any insight in the response in the least-effective modality. However, although the other criteria are based on a combination of both unisensory responses, different combinations can lead to the same threshold. This points out that no matter which criterion is used, it is of utmost importance to inspect and report the unisensory response levels (% signal change, averaged time-courses, or b-estimates) in addition to showing maps for a certain test (Beauchamp 2005b). This is necessary to fully understand why a criterion has been met by a voxel or region, and to judge the meaningfulness of a certain classification.

Because all above discussed criteria for comparing multisensory to unisensory responses have limitations, an interesting alternative is to manipulate the congruency of the different inputs (Doehrmann and Naumer 2008), for instance with regard to stimulus identity (Van Atteveldt et al. 2004) or statistical relation (Baier et al. 2006). 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. This comparison follows the assumption that a distinction between congruent and incongruent cross-modal stimulus pairs can not be established unless the unimodal inputs have been integrated successfully; therefore the congruency contrast can be used as a supplemental criterion for multisensory integration. An additional advantage of using congruency manipulations is that it facilitates inclusion of different factors within the same design, e.g., temporal, spatial and/or semantic relation between the cross-modal inputs. Such multi-factorial designs allow to directly addressing questions regarding relative contributions and interactions between different (binding) factors (Sestieri et al. 2006; Van Atteveldt et al. 2007; Blau et al. 2008; Noppeney et al. 2008). Interestingly, between-group factors can also be included in such models to assess group differences in integration, as in a recent study that revealed defective multisensory integration of speech sounds and written letters in developmental dyslexia (Blau et al. 2009).

Unsolved issues and suggested approaches

Whichever statistical criterion applied, the subsequent interpretation will always be limited because of the heterogeneity of the measured voxels. As already pointed out in the introduction, an observed voxel-level response can have many different origins at the neuronal level because voxels are likely to consist of mixed populations of uni- and multisensory neurons (Laurienti et al. 2005; Driver and Noesselt 2008). As a consequence, interpretations following the above outlined statistical criteria are never exclusive, and alternative designs and analytical tools need be explored. In the following, we will discuss alternative designs and analytical approaches that might circumvent some of the problems caused by heterogeneity within voxels (fMRI-adaptation), or make optimal use of spatially heterogeneous response patterns (MVPA). Finally, technical advancements pushing the limits of high-resolution fMRI will be considered. As we will see, resolution in the scale of cortical columns is in reach, which might drastically reduce undesired heterogeneity within the units of measurement.

Alternative fMRI design: fMRI-adaptation

The fMRI-adaptation paradigm (fMRI-A) is based on the phenomenon of reduced neural activity for repeated stimuli (repetition suppression) and hypothesizes that by targeting specific neuronal populations within voxels, their func-

tional properties can be measured beyond the voxel resolution (Grill-Spector and Malach 2001; Grill-Spector 2006). The typical procedure is to adapt a neuronal population by repeated presentation of the same stimulus in a control condition (fMRI signal reduces), and to vary one stimulus property and assess recovery from adaptation in the main condition(s). If adaptation remains (fMRI signal stays low), the adapted neurons respond invariantly to the manipulated property, whereas a recovered (i.e., increased) fMRI signal indicates sensitivity to that property, i.e., that at least partially, a different set of neurons is responding within the voxel. Within sensory systems, there are many examples in which fMRI-A revealed organizational structures that could not be revealed using more standard stimulation designs. Since (presumably) only the targeted neural population adapts, its functional properties can be investigated without being mixed with responses of other neural populations within the same voxel. In the visual system for example, heterogeneous clusters of feature-selective neurons (e.g., for different object orientations) within voxels were revealed using fMRI-A (Grill-Spector et al. 1999), whereas in a more standard stimulation design, the averaged voxel-response was not different for the different features since all of them activated a neural population within that voxel. Interestingly, fMRI-A has also been used to investigate sub-voxel level integration of features within the visual modality (Self and Zeki 2005; Sarkheil et al. 2008). Note, however, that the exact neuronal mechanism underlying BOLD adaptation is still uncertain (Grill-Spector et al. 2006; Krekelberg et al. 2006; Sawamura et al. 2006; Bartels et al. 2008).

Human “multisensory” cortex is most likely composed of a mixture of unisensory and multisensory subpopulations. In a high-resolution fMRI study, Beauchamp and colleagues demonstrated that human multisensory STS consists of mixed visual, auditory and audiovisual subpopulations (Beauchamp et al. 2004). These different neuron types were organized in clusters on a millimeter scale, which might indicate an organizational structure similar to that of cortical columns, as is also indicated by anatomical work in macaques (Seltzer et al. 1996). Cortical columns consist of about hundred thousand neurons with similar response specificity, for example, orientation columns in V1 (Hubel and Wiesel 1974), or feature-selective columns in inferotemporal visual cortex (Fujita et al. 1992). As outlined above, such a heterogeneous organization of unisensory and multisensory neuronal populations below the voxel resolution (typically around $3 \times 3 \times 3 = 27 \text{ mm}^3$), limits the certainty with which voxel-level responses can be interpreted in terms of neuronal processes. A clear example is that an enhanced BOLD response to multisensory stimulation can be due to integration at the neuronal level, but it can be explained equally well by a mix of two separate

unisensory populations. fMRI-A might be helpful to distinguish between voxels in multisensory cortex containing only unisensory neuronal subpopulations and voxels composed of a mixture of uni- and multisensory populations. Different adaptation and recovery responses could shed light on the sub-voxel organization: multisensory neurons should adapt to cross-modal repetitions (alternating modalities, e.g., A-V), while unisensory neurons should not or at least less (see below). This might be used to disentangle unisensory and multisensory neural populations. Another approach is to present repetitions of multisensory stimuli and vary the (semantic or other) relation between them (e.g., congruent vs. incongruent pairs), to test whether or not voxels contain neurons that are sensitive to this relation, assuming that these should be multisensory (Van Atteveldt et al. 2008).

Unfortunately, there are several potential pitfalls for such designs. Whereas in the example from the visual system, different neuronal subpopulations are selective to one-specific feature (e.g., maximum response to an orientation of 30°) or are not selective (e.g., orientation-invariant); different populations in multisensory cortex can be selective to “features” in different conditions: visual repetitions may adapt visual and audiovisual neurons, auditory repetitions may adapt auditory and audiovisual neurons. This can be problematic because neurons are shown to adapt despite intervening stimuli (Grill-Spector 2006), so stimulus repetitions in alternating modalities will also adapt unisensory neurons (although probably to a weaker extent). Another problem is that a cross-modal repetition (e.g., visual–auditory) may suppress activity of multisensory neurons, but will also activate new pools of unisensory neurons (in this example: auditory) in the same voxel with mixed neuronal populations, which may counteract the cross-modal suppression. In sum, fMRI-A designs to investigate multisensory integration may help interpreting representational coding at the neuronal level, but great caution is warranted.

Alternative analytical approach: multivariate statistics

While standard hypothesis-driven fMRI analyses using the GLM process the time-course of each voxel independently, and data-driven methods, such as independent component analysis search for functional networks in the whole four-dimensional data set, several new analysis approaches focus on rather local, MVPA methods (Haxby et al. 2001; Haynes and Rees 2005; Kamitani and Tong 2005; Kriegeskorte et al. 2006; De Martino et al. 2008) In these approaches, data from individual voxels within a region are jointly analyzed. An activity pattern is represented as a feature vector where each feature refers to an estimated response measure of a specific voxel. The dimension N of

an fMRI feature vector, thus, corresponds to the number of included voxels in the analysis. Using standard statistical tools, distributed activity patterns corresponding to different conditions may be compared using multivariate approaches (e.g., MANOVA). Alternatively, machine learning tools [(e.g., support vector machines (SVMs))] are trained on a subset of the data while leaving some data aside for testing generalization performance of the trained “machine” (classifier). More robust estimates of the generalization performance of a classifier are obtained by cross-validation techniques involving multiple splits of the data in training and test sets. To solve difficult classification problems, non-linear classifiers may be used but for problems with high-dimensional feature vectors and a relatively small number of training patterns, non-linear kernels are usually not required. This is important for fMRI applications, because only linear classifiers allow to use the obtained weight values (one per voxel) for direct visualization of the voxels’ contribution to the classification performance. Linear classifiers, thus, allow to perform “multivariate brain mapping” by localizing discriminative voxels. By properly weighting the contribution of individual voxels across a region (local), multivariate pattern analyses approaches open the possibility to separate spatially distributed patterns, which would be potentially too weak to be discovered by univariate (voxel wise) analysis. Note that the joint analysis of weak signals from multiple voxels does not require that voxels within a region behave in the same way, since it extracts discriminative information within the multivariate signal. If, for example, some voxels in a local neighborhood show a weak increase and other voxels a weak decrease when comparing two conditions, these opposing effects would cancel out in a regional average using a standard GLM analysis, but would well contribute to a measure of multivariate information. The gained higher sensitivity allows to separate similar distributed representations from each other due to the integration of weak information differences across voxels. Note that a high sensitivity for distinguishing distributed representations would be important even if a columnar-level resolution would allow a more direct mapping of representational units since representations of exemplars (e.g., two faces) might only slightly differ in their distributed code across the same basic features.

A recent publication (Formisano et al. 2008) has shown that distributed patterns extending from early auditory cortex into STS/STG contain enough information to reliably separate responses evoked by individual vowels spoken by different speakers from each other; performance of trained classifiers was indicated by successful generalization to new exemplars of learned vowels even if they were spoken by a novel speaker. Note that this discriminative power could not be observed when analyzing responses from single voxels. Another relevant recent fMRI study

demonstrated that classifiers (SVMs) that were trained to separate sensory events using activation patterns in premotor cortex, could also reliably separate the corresponding motor events (Etzet et al. 2008).

We propose to follow a similar approach to investigate the multisensory nature of sensory-sensory representations in multisensory cortex. It would, for example, be interesting to teach classifiers to discriminate responses from multisensory (e.g., audiovisual) stimuli in order to obtain more information about how specific stimulus combinations are represented in sensory-specific cortex (e.g., visual and auditory association cortex) and multisensory cortex (e.g., STS); different training signals (classification labels) could be used for the same fMRI data by either learning labels representing the full cross-modal pairs or by using only the visual or auditory component of a pair. These different training tasks should reveal which parts in the cortical network would more prominently code for visual, auditory or the combination of both stimuli; furthermore, if learning and generalization would be successful, the identified representations would allow predicting from a single trial response which specific audio-visual combination was presented to the subject. Such knowledge would be highly relevant for building computational models of multisensory processing, which will be discussed in the final section.

Increased spatial resolution

Increasing spatial resolution is an obvious approach to obtain more detailed fMRI data, which might additionally help to shed some light on the fine-grained functional organization of small areas in the human brain (Logothetis 2008). 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 Tesla fMRI study showed tonotopic, mirror-symmetric maps within early human auditory cortex (Formisano et al. 2003). The described tonotopic maps were much smaller than the large retinotopic maps in human visual cortex, which could therefore be observed already 10 years earlier with 1.5 Tesla scanners (Sereno et al. 1995). Another example is the study of Beauchamp and colleagues (2004), in which the authors used parallel imaging to achieve a spatial resolution of $1.6 \times 1.6 \times 1.6 \text{ mm}^3$, providing insight in the more detailed organization of uni- and multisensory clusters in posterior STS.

Despite progress in high-resolution functional imaging, it is unclear what level of effective spatial resolution can be achieved with fMRI since the ultimate spatial (and temporal) resolution of fMRI is not primarily limited by technical constraints but by properties of the vascular system. The

spatial resolution of the vascular system, and hence fMRI, seems to be in the order of 1 millimeter since relevant blood vessels run vertically through cortex in a distance of about a millimeter (Duvernoy et al. 1981). An achievable resolution of 0.5–1 mm might be just enough to resolve cortical columns (Mountcastle 1997). According to theoretical reasoning and empirical data (e.g., orientation columns in V1; Hubel and Wiesel 1974), a cortical column is assumed to contain about hundred thousand neurons with similar response specificity. A conventional brain area, such as the fusiform face area, could contain a set of about 100 cortical columns, each coding a different elementary (e.g., face) feature. Cortical columns could, thus, form the basic building blocks of complex distributed representations (Fujita et al. 1992). Different entities within a specific area would be coded by a specific distributed activity pattern across cortical columns, like letters or vowels in superior temporal cortex (Formisano et al. 2008). If this reasoning is correct, an important research strategy would aim to unravel the specific building blocks in various parts of the cortex, including individual representations of letters, speech sounds and their combinations in auditory cortex and heteromodal areas in the STS/STG. Since neurons within a cortical column code for roughly the same feature, measuring the brain at the level of cortical columns promises to provide a relevant level for revealing meaningful distributed neuronal codes. Recently it became indeed possible to reliably measure orientation columns in the human primary visual cortex using high-field (7 Tesla) fMRI (Yacoub et al. 2008). This study clearly indicates that columnar-resolution fMRI is indeed possible—at least when using high-field fMRI combined with spin-echo MRI pulse sequences. If the organization of uni- and multisensory neuronal populations is organized at a columnar level, which is hinted at by fMRI (Beauchamp et al. 2004) and animal work (Seltzer et al. 1996), this would strongly increase the feasibility of high-field fMRI to provide insight in neuronal multisensory integration because putative multisensory and unisensory columns could separately be measured, and thus distributed activity patterns across them could be analyzed.

Making optimal use of spatial resolution in group analyses: cortex-based alignment

Inspection and reporting of individual activation effects is very important, but some effects may only reach significance when performing group analyses. Moreover, random-effects group analysis are essential for assessing consistency of effects within groups, and to reveal differences between groups. Unfortunately the typically used coarse brain normalization in volume space (e.g., Talairach or MNI space) compromises the gain of high-resolution

imaging since sufficient spatial correspondence can only be achieved through substantial spatial smoothing (e.g., with a Gaussian kernel with a FWHM of 8–12 mm). Since many interesting multisensory effects may be observed only in fine-grained activity patterns, spatial smoothing should be minimal or completely avoided. Moreover, standard volumetric Talairach or MNI template brain matching techniques may lead to suboptimal multi-subject results due to poor spatial correspondence of relevant areas (Van Essen and Dierker 2007). Surface-based techniques aligning gyri and sulci across subjects (Fischl et al. 1999; Van Atteveldt et al. 2004; Goebel et al. 2006) may substantially improve spatial correspondence between homolog macro-anatomical brain structures such as STS/STG across subjects. Such an improved alignment may provide more sensitive statistical results under the assumption that functional regions “respect” macro-anatomical landmarks (Spiridon et al. 2005; Hinds et al. 2009); a systematic and large-scale functional-anatomical correspondence project is currently in progress to verify this assumption (Frost and Goebel 2008).

The effectiveness of surface-based alignment procedures on group statistical maps have been reported in several recent studies (reviewed in Van Essen and Dierker 2007). Here, we show its effectiveness also for multisensory cortical areas, such as those demonstrated in STS/STG. Figure 2 shows a direct comparison of surface-based and volume-based (Talairach) registration of group data for a multisensory investigation of letter-sound integration (Van Atteveldt et al. 2004). The figure illustrates that the analysis with cortex-based aligned data improved the statistics and provided more accurate localization of multisensory effects in auditory cortex and STS. In Fig. 2a, the max criterion resulted in a more robust map (higher threshold) that is much more clearly localized on STS (and additional clusters on STG) using cortex-based alignment. It is important to verify this statement with regard to localization by comparing the group maps with individual localizations. The details of individual STS ROIs (reported in Van Atteveldt et al. 2004) show variability in Talairach coordinates: average \pm standard deviation (x, y, z) = $(-54 \pm 4, -33 \pm 11, 7 \pm 6)$, so most variability in the y coordinate (=anterior–posterior axis). These individual ROIs were selected based on individual anatomy, i.e., they were all located on the STS. Importantly, comparison of the Talairach and cortex-based group statistical maps (Fig. 2a) indicates that the averaged Talairach coordinates do not correspond to the location of the individual ROIs on STS (Fig. 2a, left: cluster is located on STG overlapping partly with auditory cortex), whereas the cortex-based aligned map shows the dominant cluster clearly localized on STS (Fig. 2a, right). Figure 2b shows that despite a variable individual anatomy of the auditory cortex indicated in the top row (5 different subjects), the cortex-based aligned group maps accurately locate the

multisensory congruency effect on Heschl’s sulcus and Planum Temporale.

As functional-anatomical correspondence may vary for different brain regions and functions, a complementary approach to account for individual variability is the use of functional localizers, which allow to “functionally align” brains (Saxe et al. 2006). Future multisensory fMRI studies could use this approach to functionally localize integration areas, e.g., by using the max criterion. Group statistics can subsequently be performed using each subject’s fMRI time-series from the functionally defined ROI in that subject. Note, however, that there are also pitfalls regarding the use of (separate) functional localizers (Friston and Henson 2006; Friston et al. 2006) and intra-subject consistency of certain localizers was recently reported to be very low (Duncan et al. 2009). Experiments incorporating functional localizers should therefore be designed with care, for instance, it might be best to embed localizer contrasts in factorial designs (i.e., orthogonal to the main manipulation of interest; Friston et al. 2006).

Dynamic processes and neuronal interactions

Several valuable approaches exist to study temporal characteristics of neural processing from fMRI time-series. For example, Granger Causality Mapping (GCM, Goebel et al. 2003; Roebroeck et al. 2005) is an effective connectivity tool with the potential to estimate the direction of influences between brain areas directly from the voxel’s time-course data, which we recently applied to assess influences to/from STS during letter-sound integration (Van Atteveldt et al. 2009). Another dynamic effective connectivity tool is dynamic causal modeling (DCM; Friston et al. 2003), which was recently used to investigate the neural mechanism of visuo-auditory incongruency effects for objects and speech (Noppeney et al. 2008). Furthermore, information about the onset of the fMRI response (BOLD latency mapping; Formisano and Goebel 2003) can reveal insight in the temporal sequence of neural events, and has successfully been applied to multisensory research (Martuzzi et al. 2007; Fuhrmann Alpert et al. 2008).

Still, using fMRI alone it is difficult to learn about fast dynamic processes in “real-time” since successive temporal events lead to an integrated BOLD response. In multisensory research, this severely limits the certainty with which cross-modal effects observed with fMRI can be interpreted in terms of processing stage (early vs. late, or feedforward vs. feedback), which is therefore often a matter of heavy debate. A prevailing example concerns the modulation of auditory cortex activity by visual speech cues (Calvert et al. 1997, 1999; Paulesu et al. 2003; Pekkola et al. 2005; reviewed in Campbell 2008), which is typically interpreted

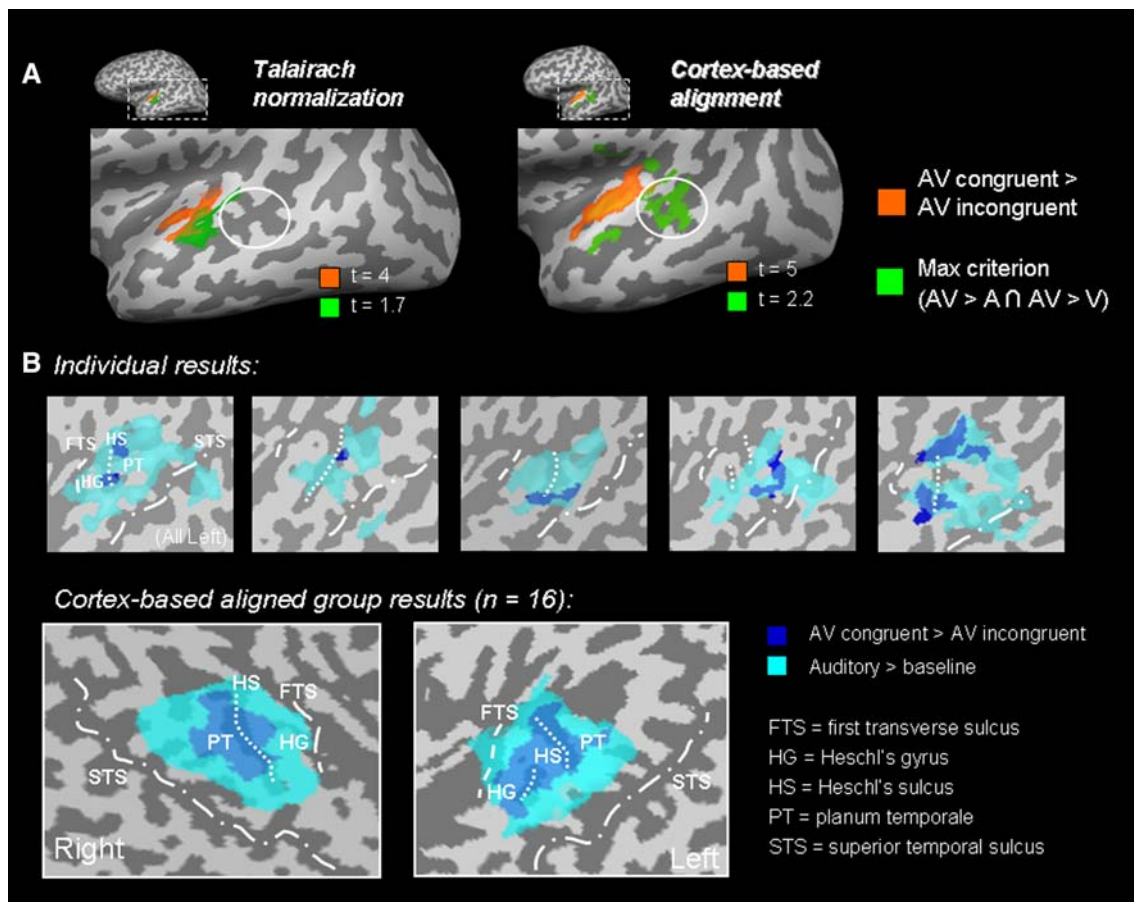


Fig. 2 fMRI group analysis results using volumetric normalization (Talairach space) and cortex-based alignment. **a** Random-effects statistical maps of two different contrasts: audiovisual congruent versus audiovisual incongruent (orange) and the max criterion expressed as the conjunction (intersection) of audiovisual versus auditory & audiovisual versus visual (green). The maps show that at higher t values, the cortex-based aligned data still provide a better group map (i.e., location of the clusters correspond best to the activations in individual subjects). **b** Individual (top row) and group (bottom row) statistical maps of the

contrasts audiovisual congruent versus audiovisual incongruent (dark blue) and auditory versus baseline (light blue). Top row shows the reconstructed and flattened cortical sheets of the left temporal lobe in five representative individual subjects, the bottom row shows the cortex-based aligned group statistics (of 16 subjects) on a representative left and right temporal lobe. White lines indicate the different sulci and borders between the gyri, from anterior to posterior: FTS first transverse sulcus, HG Heschl's gyrus, HS Heschl's sulcus, PT planum temporale, STS superior temporal sulcus

as being the result of feedback projections from the heteromodal STS/STG (reviewed in Calvert 2001). As stated in the introduction, because of its whole-brain coverage and non-invasive nature, fMRI can be used to raise specific new predictions that can be tested by intracranial human recordings, which have both spatial and temporal resolution. This has been done recently for the case of audiovisual speech processing. Besle and colleagues (2008) recorded ERP's intracranially from precise locations in the temporal lobe during audiovisual speech processing, and demonstrated that visual influences in (secondary) auditory cortex occurred earlier in time than the effects in STS/STG. These findings advocate a direct feedforward activation of auditory cortex by visual speech information.

Another very promising direction is to directly integrate fMRI and EEG/MEG. While progress has been made in recent years (Lin et al. 2006; Goebel and Esposito

2009), it remains difficult to reliably separate closely spaced electrical sources from each other. This is of relevance for multisensory research in case nearby located auditory and multisensory superior temporal regions need to be separated. The combination of EEG and fMRI data sets into one unique data model is still a focus of intensive research and requires enormous efforts to integrate independent fields of knowledge such as physics, computer science and neuroscience. In fact, besides the classical problems of head modeling in EEG, and hemodynamic modeling in fMRI, an additional difficulty is given by the need of understanding and modeling the ongoing correlations of EEG and fMRI data. Some of these problems are solved by direct intra-cranial electrical recordings from the human brain (e.g., inverse modeling), but such studies are limited because they are invasive and depend on patient populations. Despite the difficulties in properly

integrating detailed temporal information from (simultaneously) recorded EEG signals into fMRI analysis, the expected insights will be essential to build advanced spatio-temporal models of multisensory processing in the human brain.

Toward computational modeling of multisensory processing

Data from a series of fMRI experiments from our group have provided insight in the likely role of brain areas involved in letter-speech sound integration as well as the information flow between these areas (reviewed in Van Atteveldt et al. 2009). As has been highlighted in the previous sections, data might soon be available providing further constraints at the representational level of individual visual, auditory and audiovisual entities such as letters, speech sounds and letter-sound combinations. In light of the richness of present and potential future multisensory data, it seems to become feasible to build computational process models to further stimulate discussions of neuronal mechanisms. In the future, we aim to implement large-scale recurrent neural network models because they (1) allow to clearly specify structural assumptions in the connection patterns within and between simulated brain areas and (2) allow to precisely predict the implications of structural assumptions by feeding the networks with relevant unimodal and bimodal stimuli (e.g., visual letters, speech sounds and their audiovisual combinations). Running spatio-temporal simulations may help to understand how “emergent” phenomena, such as audiovisual congruency effects in auditory cortex (Van Atteveldt et al. 2004), result from multiple simultaneously operating synaptic influences from different modeled brain regions. Such neural models may also help to link results from electrophysiological animal studies and fMRI studies. The fMRI BOLD data reflect a mix of (suprathreshold) spiking activity, hemodynamic spread, and neural spread of subthreshold neural activity (Logothetis et al. 2001; Logothetis and Wandell 2004; Oeltermann et al. 2007; Maier et al. 2008). To investigate discrepancies between spiking, LFP and BOLD data, these signals must be modeled separately in an environment that permits a comparison with matching empirical data (Goebel and De Weerd 2009). To compare activity patterns in simulated neural networks with empirical data, modeled cortical columns can be linked to topographically matching voxels; these links implement spatial hypotheses and are obtained from structural brain scans and functional mapping studies in human subjects, thereby establishing a common representational space for simulated and measured data.

This allows to “run” large-scale neural network models “in the brain” and to analyze predicted fMRI data using the same analysis tools as used for the measured data (e.g., GLM, MVPA, GCM). Such a tight integration of computational modeling and fMRI data may help to test and compare the implications of specified neuronal coding principles and to study the evolution of dynamic interactions (Goebel and De Weerd 2009). If these principles are applied to multisensory experiments, assumptions about the proportion of unisensory and multisensory neurons in voxels in different brain areas can be explicitly explored and derived predictions can be tested by conducting theory-guided neuroimaging studies.

Conclusion

When using fMRI at standard resolution ($\leq 3T$, single-coil imaging) and (mass-) univariate statistical analysis (e.g., using the GLM), the statistical criteria for “multisensory integration” should be selected with care, and response characteristics of all uni- and multisensory conditions should always be inspected and reported. Alternative designs such as fMRI-adaptation may provide additional insights in multisensory integration beyond the voxel level. In addition to analyzing single-voxel responses, MVPA is an important new statistical approach to jointly analyze locally distributed activity patterns of multiple voxels. Because of the putative heterogeneous organization of multisensory brain areas, distinct integrative states may be expressed in such distributed activity patterns rather than in the response of separate voxels. Increased spatial resolution might ultimately lead to columnar-level resolution. If multisensory cortex is organized at the columnar level, fMRI might be able to separate uni- and multisensory responses, making the application of MVPA even more interesting. When data are aligned based on individual cortical anatomy, the high spatial resolution of fMRI can also be fully exploited at the group-level. Although also profiting from higher field strength and imaging technology, the ultimate temporal resolution is limited by the vascular system (and neuro-vascular coupling) and will not be sufficient to capture fast dynamic neural processes and interactions. Additional information from complementary imaging modalities (EEG, MEG) should therefore be acquired and integrated. Furthermore, direct comparison (in a common representational space) of modeled neural activity and the corresponding BOLD activity (predicted fMRI data) with real fMRI data will help interpreting multisensory hemodynamic data in terms of neural mechanisms.

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

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Amedi A, von Kriegstein K, Van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571
- Baier B, Kleinschmidt A, Muller NG (2006) Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *J Neurosci* 26:12260–12265
- Beauchamp M (2005a) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol* 15:1–9
- Beauchamp M (2005b) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113
- Beauchamp MS, Argall BD, Bodurka J, Duyn J, Martin A (2004) Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190–1192
- Besle J, Fischer C, Bidet-Caulet A, Lecaigard F, Bertrand O, Giard MH (2008) Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *J Neurosci* 28:14301–14310
- Blau V, Van Atteveldt N, Formisano E, Goebel R, Blomert L (2008) Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur J Neurosci* 28:500–509
- 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. *Curr Biol* 19:503–508
- Buxton RB, Uludağ K, Dubowitz DJ, Liu TT (2004) Modeling the hemodynamic response to brain activation. *Neuroimage* 23:S220–S233
- Bartels A, Logothetis NK, Moutoussis K (2008) fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends Neurosci* 31(9):444–453
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–596
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci* 363:1001–1010
- Cappe C, Morel A, Barone P, Rouiller EM (2009) The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cerebral Cortex*, Jan 15 [Epub ahead of print]
- 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
- Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res* 1242:136–150
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* 57:11–23
- Duncan KJ, Pattamadilok C, Knierim I, Devlin JT (2009) Consistency and variability in functional localisers. *Neuroimage* 46(4):1018–1026
- Duvernoy HM, Delon S, Vannson JL (1981) Cortical blood vessels of the human brain. *Brain Res Bull* 7:519–579
- Etzel JA, Gazzola V, Keysers C (2008) Testing simulation theory with cross-modal multivariate classification of fMRI data. *PLoS ONE* 3:e3690
- Fischl B, Sereno MI, Tootel RBH, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284
- Formisano E, Goebel R (2003) Tracking cognitive processes with functional MRI mental chronometry. *Curr Opin Neurobiol* 13:174–181
- 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
- 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
- Friston KJ, Henson RN (2006) Commentary on: divide and conquer; a defence of functional localisers. *Neuroimage* 30:1097–1099
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1994) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210
- Friston KJ, Buchel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* 6:218–229
- Friston K, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19:1273–1302
- Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN (2006) A critique of functional localisers. *Neuroimage* 30:1077–1087
- Frost M, Goebel R (2008) The structural-functional correspondence project. *Neuroimage* 41
- Fuhrmann Alpert G, Hein G, Tsai N, Naumer MJ, Knight RT (2008) Temporal characteristics of audiovisual information processing. *J Neurosci* 20:5344–5349
- Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–346
- Ghazanfar AA, Schroeder C (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285
- Goebel R, De Weerd P (2009) Perceptual filling-in: from experimental data to neural network modeling. In: Gazzaniga MS (ed) *The cognitive neurosciences IV*. MIT Press, Cambridge
- 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, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn Reson Imaging* 21:1251–1261
- Goebel R, Esposito F, Formisano E (2006) Analysis of Functional Image Analysis Contest (FIAC) data with BrainVoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392–402
- Grill-Spector K (2006) Selectivity of adaptation in single units: implications for fMRI experiments. *Neuron* 49:170–171
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol* 107:293–321
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203

- Grill-Spector K, Henson R, Martin A (2006) Repetition in the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23
- Hackett TA, De La Mothe LA, Ulbert I, Karmos G, Smiley J, Schroeder CE (2007) Multisensory convergence in auditory cortex, II thalamocortical connections of the caudal superior temporal plane. *J Comp Neurol* 502:924–952
- Haller S, Wetzel SG, Radue EW, Bilecen D (2006) Mapping continuous neuronal activation without an ON–OFF paradigm: initial results of BOLD ceiling fMRI. *Eur J Neurosci* 24:2672–2678
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430
- Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691
- Hinds O, Polimeni JR, Rajendran N, Balasubramanian M, Amunts K, Zilles K, Schwartz EL, Fischl B, Triantafyllou C (2009) Locating the functional and anatomical boundaries of human primary visual cortex. *Neuroimage* 46(4):915–922
- Hubel DH, Wiesel TN (1974) Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol* 158:267–293
- Huettel SA, Song AW, McCarthy G (2004) Functional magnetic resonance imaging. Sinauer Associates, Inc, Sunderland
- Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212:121–132
- Krekelberg B, Boynton GM, van Wezel RJA (2006) Adaptation: from single cells to BOLD signals. *Trends Neurosci* 29:250–256
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289–297
- Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hämäläinen MS (2006) Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31:160–171
- Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature* 453:869–878
- Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157
- Macaluso E (2006) Multisensory processing in sensory-specific cortical areas. *Neuroscientist* 12:327–338
- Maier A, Wilke M, Aura C, Zhu C, Ye F, Leopold DA (2008) Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat Neurosci* 11:1193–1200
- Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA (2007) Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb Cortex* 17:1672–1679
- Mountcastle VB (1997) The columnar organization of the neocortex. *Brain* 120:701–722
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609
- Oeltermann A, Augath MA, Logothetis NK (2007) Simultaneous recording of neuronal signals and functional NMR imaging. *Magn Reson Imaging* 25:760–774
- Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003) A functional-anatomical model for lipreading. *J Neurophysiol* 90:2005–2013
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16:125–128
- Perrault TJJ, Vaughan JW, Stein BE, Wallace MT (2005) Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J Neurophysiol* 93:2575–2586
- Roebroeck A, Formisano E, Goebel R (2005) Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* 25:230–242
- Sarkheil P, Vuong QC, Bühlhoff HH, Noppeney U (2008) The integration of higher order form and motion by the human brain. *Neuroimage* 42:1529–1536
- Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron* 49:307–318
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30:1088–1096
- Schroeder CE, Foxe JJ (2005) Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol* 15:1–5
- Schroeder CE, Smiley JF, Fu KG, McGinnis T, O’Connell MN, Hackett TA (2003) Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int J Psychophysiol* 50:5–17
- Self M, Zeki S (2005) The integration of colour and motion by the human visual brain. *Cereb Cortex* 15:1270–1279
- Seltzer B, Cola MG, Gutierrez C, Masee M, Weldon C, Cusick CG (1996) Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. *J Comp Neurol* 370:173–190
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893
- Sestieri C, Di Matteo R, Ferretti A, Del Gratta C, Caulo M, Tartaro A, Olivetti Belardinelli M, Romania GL (2006) “What” versus “Where” in the audiovisual domain: an fMRI study. *Neuroimage* 33:672–680
- Spiridon M, Fischl B, Kanwisher N (2005) Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp* 27:77–89
- Stein BE, Meredith MA (1993) *The merging of the senses*. MIT Press, Cambridge
- Stevenson RA, Geoghegan ML, James TW (2008) Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Exp Brain Res* 179:85–95
- 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 NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962–974
- 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 NM, Roebroeck A, Goebel R (2009) Interaction of speech and script in human auditory cortex: insights from neuroimaging and effective connectivity. *Hear Res*, June 2 [Epub ahead of print]
- Van Essen DC, Dierker DL (2007) Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56:209–225
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246–1266
- Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci USA* 105:10607–10612