Multisensory Interactions within Human Primary Cortices Revealed by BOLD Dynamics

Roberto Martuzzi[1,2], Micah M. Murray[1,3], Christoph M. Michel[4], Jean-Philippe Thiran[2], Philippe P. Maeder[1], Stephanie Clarke[3] and Reto A. Meuli[1]

[1]Service de Radiodiagnostic et Radiologie Interventionnelle, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland, [2]Institut de Traitement des Signaux, Ecole Polytechnique Fédérale, Lausanne, Switzerland, [3]Division Autonome de Neuropsychologie, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland and [4]Functional Brain Mapping Laboratory, Department of Neurology, University Hospital, Geneva, Switzerland

The first 2 authors contributed equally to this study.

Whether signals from different sensory modalities converge and interact within primary cortices in humans is unresolved, despite emerging evidence in animals. This is partially because of debates concerning the appropriate analyses of functional magnetic resonance imaging (fMRI) data in response to multisensory phenomena. Using event-related fMRI, we observed that simple auditory stimuli (noise bursts) activated primary visual cortices and that simple visual stimuli (checkerboards) activated primary auditory cortices, indicative of multisensory convergence. Moreover, analyses of blood oxygen level–dependent response dynamics revealed facilitation of hemodynamic response peak latencies and slopes for multisensory auditory–visual stimuli versus either unisensory condition, indicative of multisensory interactions within primary sensory cortices. Neural processing at the lowest cortical levels can be modulated by interactions between the senses. Temporal information in fMRI data can reveal these modulations and overcome analytic and interpretational challenges of more traditional procedures. In addition to providing an essential translational link with animal models, these results suggest that longstanding notions of cortical organization need to be revised to include multisensory interactions as an inherent component of functional brain organization.

Keywords: auditory, cross-modal, fMRI, multisensory integration, primary cortex, visual

Introduction

Sensory inputs converge and interact, influencing perception and behavior (e.g. Stein and Meredith 1993). Neurophysiological bases for these multisensory effects are increasingly being investigated. Anatomical studies in animals have identified direct, monosynaptic projections between primary and immediately adjacent auditory cortices and primary visual cortices (Falchier and others 2002; Rockland and Ojima 2003; Clavagnier and others 2004). Electrophysiological recordings showed multisensory effects within primary and adjacent auditory regions of monkeys (e.g. Ghazanfar and others 2005; Schroeder and Foxe 2005) and nonlinear response interactions within the initial 100 ms poststimulus onset in humans (e.g. Giard and Peronnet 1999; Foxe and others 2000; Molholm and others 2002; Gonzalez Andino and others 2005; Murray and others 2005). The earliest temporal stages of cortical processing and brain areas traditionally held to be unisensory in their function thus exhibit multisensory interactions.

Regarding interactions between the auditory and visual systems, several questions remain unresolved. For example, it is unknown whether monosynaptic projections between primary cortices in monkeys also exist in humans and, if so, whether they produce multisensory interactions that are measurable noninvasively. Although the former is currently not feasible with existing tracing methods, the latter can be assessed using brain imaging techniques. Convergence and interaction effects have been obtained for speech/faces (e.g. Calvert 2001; Pekkola and others 2005), letters/vocalizations (e.g. van Atteveldt and others 2004), and environmental objects (e.g. Beauchamp and others 2004).

However, the use of functional magnetic resonance imaging (fMRI) to identify multisensory effects is debated (Calvert 2001; Beauchamp 2005; Laurienti and others 2005). In particular, it is unclear whether criteria that have been applied in electrophysiological studies at the single neuron level are valid for fMRI analyses. For example, criteria for convergence (i.e. responding to multiple senses), multisensory enhancement (i.e. responding more to multisensory than to both unisensory stimuli), supra-additivity (i.e. responding more to multisensory than to the summed unisensory responses), and sensitivity to congruent stimulus features (e.g. spatial position, temporal coincidence, or object-related/semantic attributes) have been argued as prone to reporting falsely positive results (Laurienti and others 2005). Despite these considerations, some studies have reported multisensory effects within primary and adjacent auditory cortices in response to somatosensory (Foxe and others 2002) or visual stimuli (e.g. Pekkola and others 2005). To date, none have observed effects within primary visual cortex or examined whether multisensory effects in low-level cortical regions can similarly be elicited by meaningless, rudimentary stimuli. Instead, investigations have thus far been limited to meaningful stimuli, and it remains unknown whether effects in low-level cortices are mediated by higher-order processes. Resolving such questions is critical for determining whether interaction mechanisms between sensory systems are a general, perhaps automatic, property or are instead regulated by stimulus-specific processes.

Interpretational caveats of standard fMRI were bypassed by analyzing dynamics of the blood oxygenation level–dependent (BOLD) signal. Recent developments in event-related fMRI indicate that latency analyses can be performed on the directly measured BOLD signal (e.g. Josephs and others 1997; Henson and others 2002; Bellgowan and others 2003; Martuzzi and others 2006). As such, we analyzed both the spatial and the temporal pattern of responses during auditory–visual (AV) integration.

Materials and Methods

Subjects

Twelve healthy subjects (mean ± SD age = 29.4 ± 7.1 years; 6 female) with normal hearing, normal or corrected-to-normal vision, and no history of neurological or psychiatric disease participated. Each provided written informed consent to procedures approved by the Ethics Committee of the Faculty of Biology and Medicine at the University of Lausanne.

Stimuli and Task

Subjects performed a simple reaction time task to visual (a non-alternating yellow on black centrally presented checkerboard, measuring $24° \times 32°$ in total size and each square covering $0.8° \times 0.8°$ of visual angle), auditory (a binaural noise burst), or simultaneous AV stimuli (each 150-ms duration). Stimulus conditions were pseudorandomly intermixed across trials. Subjects were instructed to respond as fast as possible with their right hand upon detection of any stimulus by pressing keys. Behavioral data were acquired using button presses on a MRI-compatible device (Photon Control Inc., Burnaby, BC, Canada). Subjects pressed 4 keys, one per finger, in one swift and continuous movement like tapping one's fingers on a table, and reaction times were recorded as the latency at which the first of the keys was pressed. Stimulus delivery and the acquisition of behavioral data were controlled by E-prime (Psychology Software Tools, Inc., Pittsburgh, PA). Behavioral data from 2 of the twelve subjects were lost due to technical failures.

The interstimulus interval varied pseudorandomly from 14.2 to 17.8 s in steps of 200 ms, allowing the BOLD signal to return to baseline between stimulus presentations (e.g. Josephs and others 1997). There was a pseudorandom, variable delay between stimulus onset time and volume acquisition of 0–1.8 s at steps of 200 ms, yielding a total of 10 different delays. Jittering stimulus presentation relative to volume acquisition permitted the BOLD response to be effectively sampled with a temporal resolution of 200 ms (see Supplementary Animation and Martuzzi and others 2006 for additional details). The experiment consisted of 4 sessions, each including 10 repetitions per experimental condition. Therefore, each of the 3 experimental conditions collectively included 4 volume acquisitions at each of the 10 delays used.

Magnetic Resonance Imaging

fMRI data were acquired using an event-related design on a 3.0-T Philips Intera system equipped with an 8-channel head coil. BOLD signals were obtained with a single shot gradient-echo echo-planer imaging sequence (repetition time [TR] = 2 s, echo time [TE] = 30 ms, field of view [FoV] = 224 mm, flip angle = 90°, matrix size 64×64). Each volume was comprised of 16 slices (slice thickness 5 mm, gap 1 mm) covering the entire cerebral hemispheres and acquired in ascending order (i.e. first slice at the bottom of the head). To provide precise structural and anatomical localization of brain activity, a sagittal T_1-weighted 3D gradient-echo sequence was acquired for each subject (160 contiguous sagittal slices, slice thickness 1 mm, matrix size 256 × 256, TR = 9.9 ms, TE = 4.6 ms, FoV = 256 mm, flip angle = 8°).

Spatial fMRI Analyses

Two types of fMRI analyses were conducted in order to investigate multisensory interactions: the first in terms of the spatial pattern of activated brain regions, the second in terms of changes in temporal dynamics within activated brain regions. The latter analysis was done by measuring the shift in peak latency of the estimated hemodynamic response function within a given area across stimulus conditions (detailed below).

Activation maps were obtained using SPM2 software (Wellcome Department of Cognitive Neurology, London, UK). Functional volumes were first spatially realigned to the first volume acquired and temporarily realigned to the first slice acquired. Volumes were then normalized to the Montreal Neurological Institute (MNI) template, resampled to a voxel size of $3 \times 3 \times 3$ mm^3, and smoothed with an isotropic Gaussian kernel (full width half maximum = 6 mm). For each subject, a high-pass filter was applied on the time series to minimize possible effects of baseline drift. The statistical analysis was performed with the General Linear Model, using as a basis function the canonical hemodynamic response function and its temporal derivative, as defined in SPM2.

Structural and functional volumes were coregistered within the same coordinate system by normalizing structural images to the MNI template brain and resampling voxels to a $1 \times 1 \times 1$ mm^3 size. Inference on the population (group analysis) was obtained by means of second-level statistics, according to the random effects theory. Analyses were conducted to determine the active regions in each condition, separately (voxel-level threshold at $P < 0.001$, uncorrected; 15 voxel spatial-extent threshold) both on a single subject basis and on the group results.

Temporal fMRI Analyses

The hemodynamic response was reconstructed by averaging the 4 samples collected at each time point relative to stimulus onset and then filtering to remove high-frequency noise. Peak latency and intensity of the hemodynamic signals were measured after fitting the signal around the peak (±0.8 s, equal to 4 data points before and after the peak) with a cubic curve. We would emphasize that cubic fitting was solely applied to these data points encompassing the peak of the acquired time course response, and that time courses shown always display the filtered raw data.

For each experimental condition, peak latencies were derived for each voxel within the brain. To ensure a valid measure of peak latency, analyses were spatially restricted to the regions identified as responsive to auditory and visual stimuli. For each subject, visual-responsive regions (VRRs) were defined according to the overlap between activation maps from the visual and multisensory conditions and auditory-responsive regions (ARRs) by the overlap between activation maps from the auditory and multisensory conditions (see Beauchamp 2005 for a similar approach in analyses of BOLD amplitude). This overlap criterion also minimizes the likelihood of falsely considering a voxel as responsive to either visual or auditory stimulation. Within VRRs we statistically tested (paired t-test) the difference between BOLD peak latencies from the visual and multisensory conditions, whereas within ARRs this analysis was between auditory and multisensory conditions. It should be noted that these tests were conducted on a voxel-wise level and that not all subjects necessarily exhibited the same VRRs and ARRs. In such instances, no peak latency would be measured for a particular subject at a specific voxel. This would result in an empty cell in the analysis matrix and thus increases the propensity for falsely negative results. This constituted the group-level analysis of shifts in BOLD peak latency. We considered only those clusters meeting both a $P < 0.05$ alpha criterion (uncorrected) and also a 15-voxel spatial-extent criterion.

Although the above approach can identify modulation in BOLD peak latency throughout the entire brain, it does not account for intersubject variability in cortical functional geometry. To partially overcome this limitation without restricting our analyses to particular anatomical subdivisions (i.e. to conduct analyses throughout the entire brain volume), for each subject we identified voxels within individual VRRs and ARRs that were also within the regions defined by the aforementioned group-level analysis of shifts in BOLD peak latency. This analysis yielded a subset of 4 regions—primary visual and auditory cortices, bilaterally (see Results). Within each of these 4 individually identified regions, we calculated the mean BOLD response so as to obtain a single time curve per condition and per subject. The peaks of these time curves, within each region separately, were statistically analyzed (multivariate test and post hoc paired t-tests) across stimulus conditions. One subject did not show a reliable response (i.e. exceeding the noise level) to auditory stimulation within primary visual areas bilaterally and was therefore excluded from analyses including this condition in these regions. Peak intensity values of these time curves were similarly analyzed. Lastly, we also analyzed the slope of these curves as a post hoc supplement to our analyses of peak latency. For each curve, the slope was defined in the following manner. First, we defined the positions and intensities of the peak and the immediately preceding minimum. From these 2 points, we then fit a line within the acquired data points lying between the 20% and 90% intensity values of this range. The slope of this line was then analyzed, as above, with a multivariate test and follow-up t-tests.

Results

Behavioral data confirmed that multisensory interactions occurred. Mean reaction times were faster for the multisensory
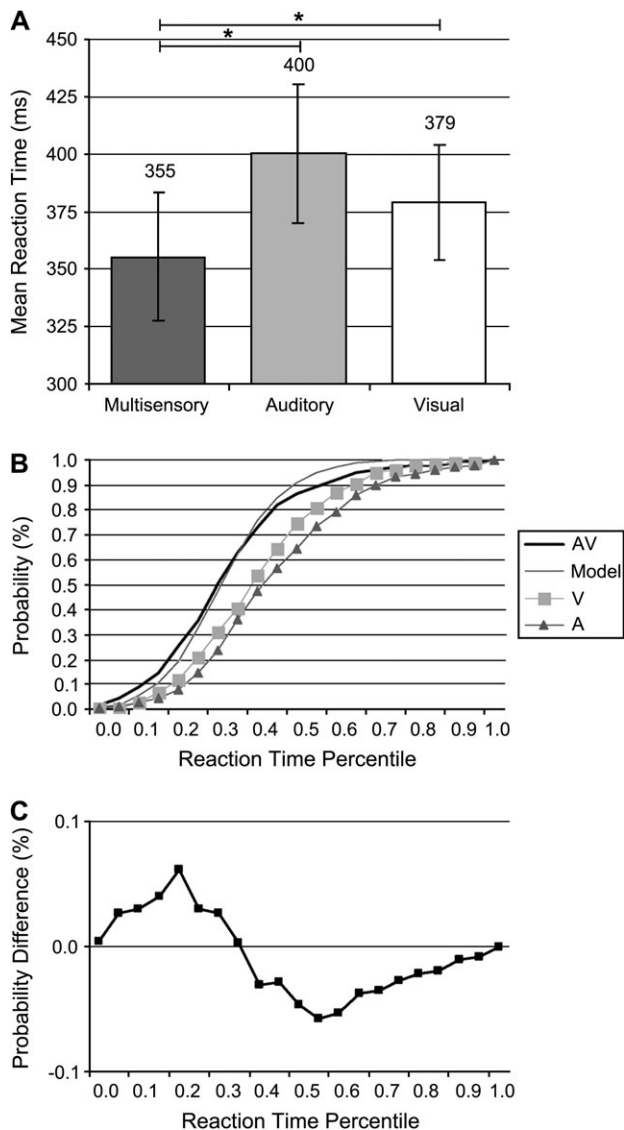
Figure 1. Mean reaction times. (a) Subjects responded faster to the multisensory than to either visual ($t_{(9)} = 3.4$; $P = 0.008$) or auditory ($t_{(9)} = 6.4$; $P < 0.001$) condition, and reaction times to unisensory stimuli did not significantly differ ($t_{(9)} = 1.8$; $P > 0.10$). Asterisk indicates significant difference ($P < 0.05$; 2-tailed paired t-test). (b) Group-average cumulative probability distributions for each stimulus condition as well as the modeled data based on Miller's (1982) inequality. (c) Reaction times to multisensory stimuli exceeded predictions of probability summation over the fastest third of the distribution (indicated by positive values).

than either visual or auditory condition (mean ± SEM = 355 ± 28 ms, 379 ± 25 ms, and 400 ± 30 ms, respectively; $F_{2,8} = 36.38$; $P < 0.001$; see Fig. 1a), replicating prior demonstrations of a redundant signals effect between audition and vision (e.g. Raab 1962; Miller 1982; Schröger and Widmann 1998; Molholm and others 2002). Additionally, this facilitation exceeded predictions from probability summation (Miller 1982; Murray and others 2005), which is a psychometric benchmark of integrative processing. Over the fastest third of the reaction time distribution, there was a higher likelihood of a reaction time following a multisensory stimulus than would be expected if auditory and visual stimuli competed independently to elicit a motor response (i.e. the so-called race model; see Miller 1982 for details; Fig. 1b,c).

Our behavioral results are suggestive of an "asymmetry" in the mean reaction time data, such that the difference between the multisensory and visual conditions is approximately the same magnitude as the difference between the visual and auditory conditions. It is important to note, however, that although the difference between the multisensory and each unisensory condition is significant, the difference between the visual and auditory conditions is not (please see caption to Fig. 1). This pattern replicates previously published studies (Schröger and Widmann 1998; Molholm and others 2002). Even if mean reaction times to the visual condition were faster than those to the auditory condition, it would be difficult to draw many direct conclusions. One reason is that we did not attempt to equilibrate the intensity of the visual and auditory stimuli, as this was not pertinent to the experimental aims. Second, as this is the first study to conduct an AV simple detection paradigm within an fMRI environment, there were no predictions as to how the scanner noise would affect performance. Third, previous research has shown that some subjects are faster with auditory stimuli and others faster with visual stimuli (Giard and Peronnet 1999). To date and to the best of our knowledge, there has been no subsequent study that would provide a solid explanation as to why this is the case (though this would certainly be an interesting avenue for future research).

Two series of event-related fMRI data analyses were conducted—one in terms of spatial activation maps, following standard procedures and another in terms of BOLD response peak latencies (see Materials and Methods). Activation maps (Fig. 2) and BOLD times series (Fig. 3) show that primary cortices of each sensory modality (i.e. calcarine cortex and Heschl's gyrus) responded to both visual and auditory stimulation, indicative of multisensory convergence. In addition, the activation maps also show robust responses within the left primary motor cortex, left somatosensory areas, the supplementary motor area (SMA), and thalamic regions. Because subjects performed a button-press to each stimulus presentation, irrespective of sensory modality, the tactile input likely elicited responses within somatosensory regions. This notion is in part supported by the left-lateralized activations in somatosensory and motor regions (see Fig. 2). Consequently, it would be difficult to disambiguate any multisensory convergence (i.e. responses to the auditory and/or visual stimuli themselves) in these areas.

In order to assess multisensory interactions (i.e. where these convergent inputs alter responses to simultaneous AV stimulation) while also minimizing issues in fMRI investigations of multisensory interactions that stem, in part, from analyses of BOLD signal amplitude (Calvert 2001; Beauchamp 2005; Laurienti and others 2005), we derived peak latencies for each brain voxel, stimulus condition, and subject, separately. These values were measured from the raw BOLD responses sampled every 200 ms. We ensured that latency measures originated from active voxels by spatially restricted temporal analyses to VRRs and ARRs, respectively (see Materials and Methods for details). Each of these paired contrasts (i.e. multisensory vs. visual and multisensory vs. auditory) revealed significant ($P < 0.05$) multisensory facilitation in terms of earlier peak BOLD response latencies principally within primary and/or near-primary visual and auditory cortices (Fig. 3a), the coordinates of which accord with ranges based on probabilistic mapping (Amunts and others 2000; Rademacher and others 2001). Although both VRRs and ARRs included other cortical and thalamic structures, no significant effects on peak latency were observed (see Supplementary Figure). Additionally, no regions showed significantly delayed multisensory responses.
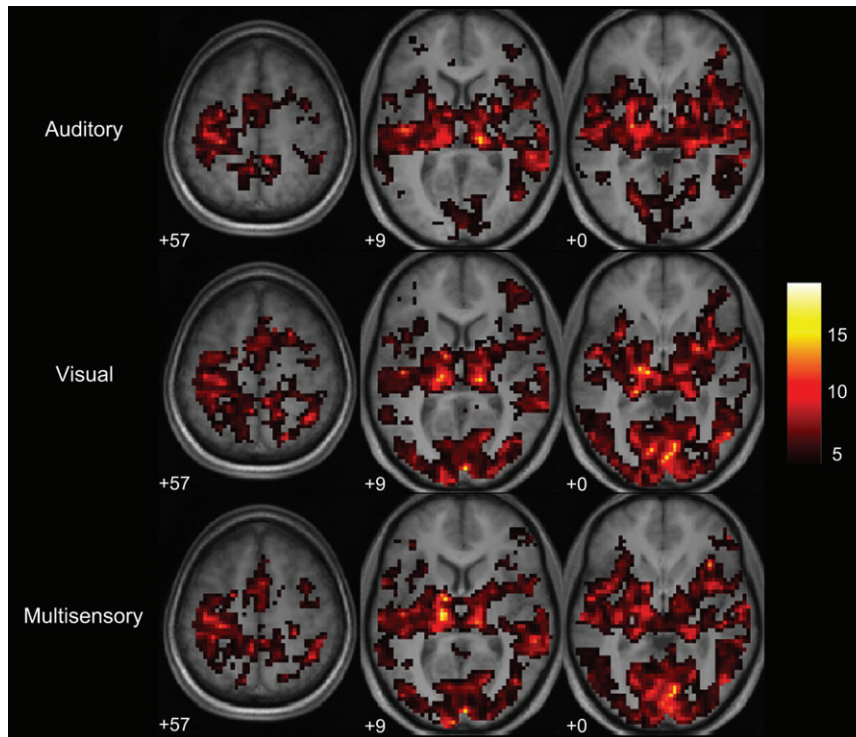
Figure 2. Activation maps ($P < 0.001$, cluster-size > 15 voxels; color scale represents t-values) for each stimulus condition show that all conditions led to responses within primary auditory and visual cortices, the left primary motor cortex, the SMA, and thalamic regions. Axial slices are shown at 3 z-coordinates (indicated in insets), using the MNI system. The left hemisphere is displayed on the left side of the image.

To this point, our analyses revealed both multisensory convergence and shifts in BOLD dynamics within primary auditory and visual cortices following simultaneous AV stimulation. This method allows us to investigate interactions throughout the entire brain but one possible shortcoming is that it does not account for intersubject anatomical and functional variability (Amunts and others 2000; Rademacher and others 2001). However, to our knowledge there is no universally accepted method for intersubject alignment of functional activations, and existing approaches minimally require the preselection of anatomical subdivisions for analyses (Pekkola and others 2005). More specific to our study, it is possible that superposition of activated areas across individuals was incomplete. That is, for this voxel-wise analysis it is possible that a given subject did not exhibit a robust BOLD response at a particular voxel, in which case no peak latency would be measured. The above analysis can therefore be considered conservative.

To partially overcome these issues, we identified voxels within individual subject VRRs and ARRs that were also within the regions defined by the aforementioned group-level analysis of BOLD peak latency shifts performed on each voxel (see Materials and Methods for details). This yielded a subset of 4 regions—primary visual and auditory cortices, bilaterally. For each subject we calculated the mean BOLD response time curve in response to each condition in each of these individually defined regions (Fig. 3b). In other words, the cluster-wise analysis used the contiguous regions defined in the above voxel-wise analyses to screen for contiguous voxels at the individual subject level. That is, we determined which voxels showed a robust BOLD response for each subject and then took the average across them before measuring the peak response latency. Importantly, this cluster-level analysis differs from the above voxel-level analysis in that each subject contributes a value to each test, thereby maintaining the degrees of freedom while also partially allowing for variation in functional anatomy. Peak latencies and intensities were statistically compared using experimental condition as the within-subjects factor. Each region showed a significant main effect of experimental condition on peak latencies that was explained by earlier peak latencies for the multisensory than either unisensory condition (Fig. 3b and upper portion of Table 1). It is important to note that the differences in peak latencies were larger than our 200 ms temporal sampling frequency, excluding the cubic fitting procedure as an explanation for the present results. A significant main effect of condition on peak intensity was also shown in each region, which was due to significantly smaller auditory responses within visual areas and smaller visual responses within auditory cortices (Fig. 3b and lower portion of Table 1). Conversely, no significant differences were obtained either between multisensory and auditory intensities within auditory areas or between multisensory and visual intensities within visual areas. Thus, the present effects on peak latency cannot follow from a simple trade-off between lower response intensity and earlier peak latencies because responses with significantly earlier peak latencies could also have significantly larger intensities (see Table 1). To further exclude such a possibility, we also assessed whether the slope of the response to the AV condition was steeper than that of either unisensory condition (see Materials and Methods for details). The results of these analyses can be found in Supplementary Table 1. In agreement with our analyses of peak latency, each region showed a significant main effect of experimental condition that was explained by a steeper slope for the multisensory than either unisensory condition (Fig. 3b and Supplementary Table 1).
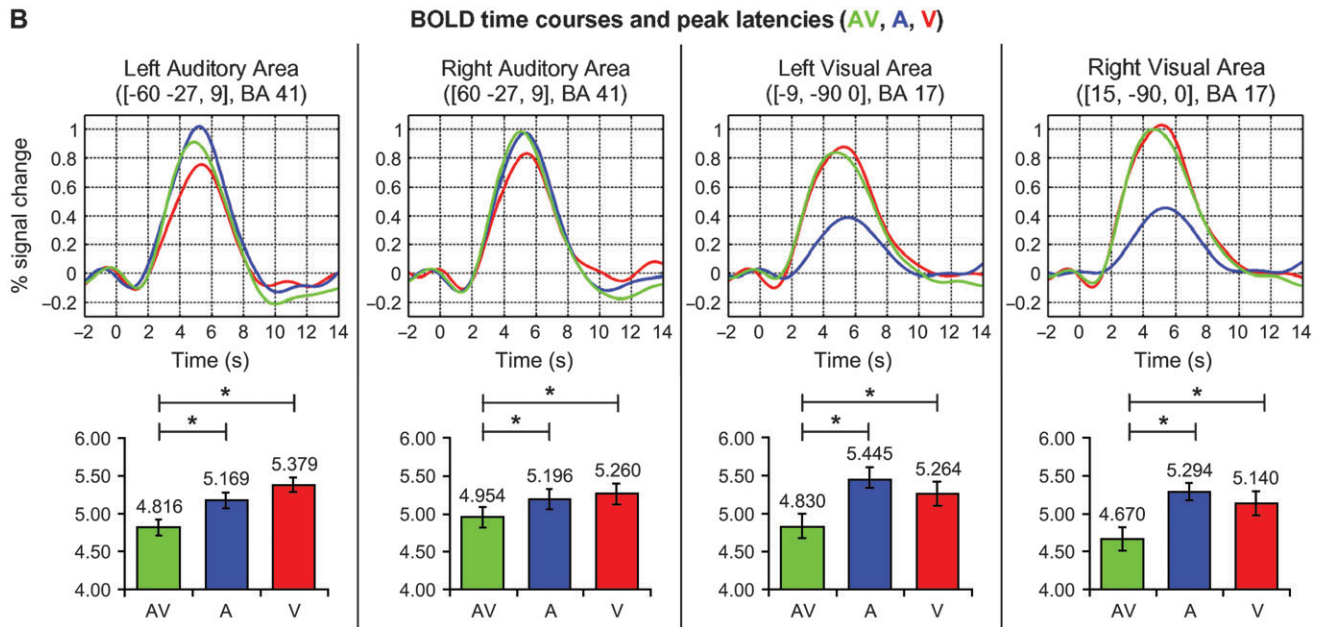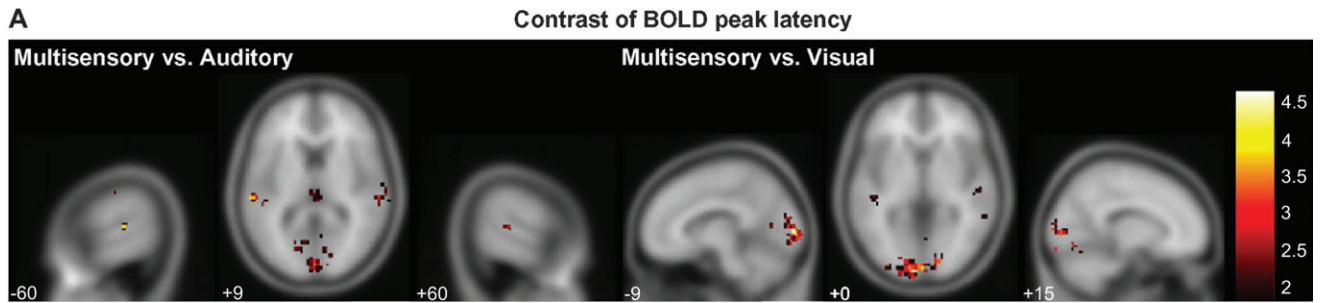
Figure 3. Facilitation in BOLD peak latencies in response to A, V, and AV multisensory stimulation (blue, red, and green traces, respectively). (a) Contrast of BOLD peak latency (paired t-test, 1-tailed, t-value indicated) at each voxel within VRRs and ARRs (see Experimental Procedures for details). (b) Dynamic shifts in BOLD peak latencies within primary cortices. MNI coordinates and Brodmann Area of the center of the cluster are indicated. Bar graphs show mean peak latencies (standard error of the mean indicated). Asterisk indicates significant difference ($P < 0.05$, paired t-test, 2-tailed).

Table 1.

Results of statistical analyses on peak latency and intensity

	Multivariate test	Follow-up comparisons (paired t-test, 2-tailed)					
		AV versus A		AV versus V		V versus A	
Peak latency							
Left auditory area	$F_{2,10} = 19.70, P = \mathbf{5.1 \cdot \times 10^{-4}}$	$t_{(11)} = -5.326$	$\mathbf{2.4 \cdot \times 10^{-4}}$	$t_{(11)} = -4.199$	**0.001**	$t_{(11)} = 1.443$	0.177
Right auditory area	$F_{2,10} = 4.941, P = \mathbf{0.032}$	$t_{(11)} = -2.512$	**0.029**	$t_{(11)} = -2.413$	**0.034**	$t_{(11)} = 0.426$	0.678
Left visual area	$F_{2,9} = 11.246, P = \mathbf{0.004}$	$t_{(10)} = -2.476$	**0.033**	$t_{(11)} = -4.417$	**0.001**	$t_{(10)} = -0.629$	0.544
Right visual area	$F_{2,9} = 4.386, P = \mathbf{0.047}$	$t_{(10)} = -2.976$	**0.014**	$t_{(11)} = -2.652$	**0.023**	$t_{(10)} = -0.746$	0.473
Peak intensity							
Left auditory area	$F_{2,10} = 4.942, P = \mathbf{0.032}$	$t_{(11)} = -1.962$	0.076	$t_{(11)} = 2.955$	**0.013**	$t_{(11)} = -3.169$	**0.009**
Right auditory area	$F_{2,10} = 7.317, P = \mathbf{0.011}$	$t_{(11)} = -0.129$	0.900	$t_{(11)} = 3.666$	**0.004**	$t_{(11)} = -2.212$	**0.049**
Left visual area	$F_{2,9} = 18.064, P = \mathbf{0.001}$	$t_{(10)} = 5.490$	$\mathbf{2.7 \cdot \times 10^{-4}}$	$t_{(11)} = -0.806$	0.437	$t_{(10)} = 6.172$	$\mathbf{1.1 \cdot \times 10^{-4}}$
Right visual area	$F_{2,9} = 27.195, P = \mathbf{1.5 \cdot \times 10^{-4}}$	$t_{(10)} = 5.500$	$\mathbf{2.6 \cdot \times 10^{-4}}$	$t_{(11)} = -0.330$	0.747	$t_{(10)} = 7.766$	$\mathbf{1.5 \cdot \times 10^{-5}}$

Note: Bold typeface indicates statistically significant values ($P < 0.05$).

As a final step, we investigated whether or not the observed shifts in BOLD peak latency correlated with the observed reaction time facilitation. To do this, we first calculated the peak latency difference between the visual (V) and AV conditions and between the auditory (A) and AV conditions for each region exhibiting a significant facilitation in BOLD peak latency for the AV condition (i.e. bilateral primary auditory and visual cortices). We also calculated the reaction time difference between the V and AV conditions and between the A and AV

conditions. In no case was a significant correlation observed (all P values > 0.05), providing no evidence for a direct correspondence between effects on BOLD peak latency and performance (see also Formisano and others 2002).

Discussion

This is the first demonstration of multisensory interactions in primary visual and auditory cortices, which manifested as

dynamic shifts in BOLD response latencies. No other regions showed significant effects on peak BOLD latency. Conjointly, we observed robust responses, in terms of BOLD amplitude, to both senses within low-level cortices. Simple visual stimuli lead to responses within auditory cortices and vice versa. These findings raise the question of their underlying neurophysiologic bases. Given the emerging anatomical and electrophysiological evidence for AV multisensory interactions within the earliest processing stages (e.g. Schroeder and Foxe 2005), our results on peak latency are most parsimoniously interpreted as consistent with direct interactions, rather than mediation by other brain regions.

A principal finding of the present study is that primary cortices responded to stimuli of other sensory systems. This is indicative of multisensory convergence, according to the criteria defined by Stein and Meredith (1993). In primary auditory cortex bilaterally, there were robust responses to visual stimuli that were nonetheless significantly smaller than the response to either auditory or multisensory stimuli (see Fig. 3 and Table 1). Similarly, in primary visual cortices bilaterally, there were responses to auditory stimuli that were significantly smaller (i.e. approximately half the magnitude) than the responses to either visual or multisensory stimuli (Fig. 3 and Table 1). Several recent studies have also documented AV convergence within primary or near-primary cortices (Calvert 2001; Pekkola and others 2005; Tanabe and others 2005). Others, using a block design, have shown deactivation in auditory cortices in response to visual stimuli and vice versa (e.g. Laurienti and others 2002; see also Haxby and others 1994; Kawashima and others 1995). However, these modulations were not present on multisensory blocks (Laurienti and others 2002), and the other studies were limited to examinations of selective attention to specific visual features. Thus, selective attention to one sensory modality might hinder the observation of multisensory effects such that positive multisensory convergence (i.e. stimuli of both senses leading to positive-going activations) may depend on paradigms that include a multisensory context and attention to multiple sensory modalities (see Laurienti and others 2002; Brosch and others 2005; Tanabe and others 2005). In the present paradigm, attention was continuously allocated to both the auditory and visual modalities because there was equal likelihood that either sense would be stimulated on a given trial.

Even if attention could account for multisensory convergence (i.e. frank responses to both sensory modalities), it cannot readily account for either the interaction effects on reaction times or the facilitation of BOLD response peak latencies for multisensory relative to both unisensory conditions in the present study (i.e. significantly faster reaction times and BOLD peak latencies to the AV condition). This is corroborated by the fact that the facilitation of reaction times (i.e. the redundant signals effect) exceeded probability summation. That is, likelihood of a fast reaction time following a multisensory stimulus was greater than the summed likelihoods of an equally fast reaction time following either unisensory stimulus (see Fig. 1). This is indicative of facilitative integrative processing exceeding any contribution of selective attention. Similarly, selective attention cannot account for the fact that BOLD peak latency was facilitated for the AV condition. If such were the case, one would expect that BOLD peak latency for the AV condition would be equivalent to the faster of the 2 unisensory conditions. Our results, however, indicate that the peak latency in response

to the AV condition was earlier than either unisensory condition and that the peak latencies in response to the 2 unisensory conditions did not significantly differ from each other (see Table 1).

Task demands, by contrast, do not appear to be a determining factor for observing multisensory convergence or interactions. For example, auditory–somatosensory multisensory convergence and supraadditive interactions have been shown using passive paradigms where subjects were nonetheless aware that stimuli would be presented in either or both sensory modalities (Foxe and others 2000, 2002). Whether or not this applies to effects on BOLD dynamics will be a topic for future experiments. However, electrophysiological studies in nonhuman primates provide one line of evidence that multisensory convergence and interactions occur under passive conditions and even under anesthesia. Frank responses to both visual and somatosensory stimuli have been recorded within primary and belt auditory cortices (e.g. Schroeder and others 2001; Fu and others 2003; Ghazanfar and others 2005; see also Kayser and others 2006 for recent fMRI results in macaques); though, to our knowledge, similar experiments within visual cortices have not yet been conducted. This collective pattern of results suggests that unisensory stimulation in a paradigm lacking a multisensory context can indeed elicit multisensory effects (in particular convergence). To date, such effects have been most consistently observed using intracranial electrophysiological methods in animals. Our results nonetheless support the future use of event-related designs in combination with high-field fMRI in noninvasively identifying multisensory phenomena under unisensory conditions.

Analyses of BOLD dynamics represent a methodological advancement for identifying multisensory brain regions with fMRI. Here, multisensory interactions led to changes in BOLD latency but not amplitude (see Fig. 3). Importantly, these latency effects did not follow from a simple amplitude/latency trade-off. In auditory cortices, responses to the multisensory condition peaked earlier than either unisensory condition, even though its amplitude was equal to that following auditory stimulation and larger than that following visual stimulation. Likewise, in visual cortices, responses to the multisensory condition peaked earlier than either unisensory condition, even though its amplitude was equal to that following visual stimulation and larger than that following auditory stimulation (see Table 1 for detailed statistics).

Several laboratories have recently been examining the validity of different analyses for identifying multisensory interactions (Calvert 2001; Beauchamp 2005; Laurienti and others 2005). As noted by one laboratory, these approaches inherently assume that signals within a given voxel emanate from a singular, homogenous neural population in terms of its responsiveness (Laurienti and others 2005). When this is not the case, canonical criteria of convergence and enhancement can be overly liberal and yield falsely positive results. Supraadditivity as an analysis criterion has also been subject to criticism. Laurienti and others (2005) contended, based largely on the frequency of observing supra- and subadditive interaction profiles at the individual neuron level, that at the level of fMRI voxels such populations would not be separable and that a supraadditive criterion would be prone to falsely negative results. Beauchamp (2005) also suggests that this criterion is overly strict. Instead, Beauchamp supports first restricting analyses to those voxels showing activation to any experimental condition and then applying

a threshold wherein the multisensory response must exceed the mean of the unisensory responses. In accord with this proposal, we first spatially restricted our analyses to voxels responsive to both audition and vision. This was done to ensure that peak latency shifts occurred in locations active under unisensory conditions, and that a peak latency shift was apparent in voxels showing a robust, positive BOLD response. More importantly, we would contend that analyses of BOLD response latency bypass the aforementioned interpretational concerns associated with analyses of BOLD response amplitude and provide a clear metric of multisensory interactions. Our methods highlight that the full range of effects may go undetected by typical analysis approaches of BOLD amplitude. One proposition is that an earlier peak reflects facilitated neural processing time (Henson and others 2002). Although appealing, it will be important for future investigations to detail more fully the bases for latency shifts in the BOLD signal.

The present data do not allow us to differentiate feedforward from feedback activity within a cortical region. Still, the anatomical studies that first identified direct projections between primary cortices noted that axon terminals were situated predominantly within layers 1 and 6, consistent with a functionally feedback profile (Rockland and Ojima 2003). This interpretation is likewise supported by electrophysiological recordings in the case of visual inputs into auditory cortices, which were distributed across the cortical laminae (Schroeder and others 2003). One speculative possibility is that the magnitude of the observed BOLD responses in the present study might be representative of the distribution of inputs into the region. In the case of visual inputs into auditory cortices, this distribution may be diffuse, whereas auditory inputs into visual cortices may be rather limited or focused. One level of support for this possibility stems from the work of Logothetis (2003) demonstrating a higher level of coupling between local field potentials, considered to be a measure of input activity within a region, and the BOLD response than between multiunit activity, considered to be a measure of the output activity within a region, and the BOLD response. Substantiating the above speculation will require further experimentation in animal models that specify the laminar origin of signals, though some work has begun in this direction (e.g. Schroeder and others 1998; Goense and Logothetis 2006). A further speculation is that direct interactions between primary auditory and visual cortices are the basis for the observed latency shifts. As mentioned above, several laboratories have now independently identified monosynaptic projections between these cortices in nonhuman primates (Falchier and others 2002; Rockland and Ojima 2003; Cappe and Barone 2005). Although such information in humans would be of immense importance, it is presently not feasible with existing staining methods. An alternative viewpoint would be that the present effects are instead mediated by another region. Although we cannot unequivocally exclude such a possibility, it is surprising that such a region did not itself show a BOLD latency shift or amplitude modulation. In addition, one might also have expected that latency shifts in primary cortices would be mirrored by effects in motor-related cortices that would in turn underlie the observed facilitation in reaction times. The present study provides no evidence that such is occurring. For one, no effects on peak latency were observed in M1, the SMA, or subcortical regions that were nonetheless identified as active under all stimulation conditions (see Supplementary Figure). Second, there was no evidence of a significant correlation between shifts in BOLD peak latency and behavioral facilitation. Finally, as discussed above, the evidence for multisensory interactions during both passive and active conditions suggests that task performance is not directly linked with interactions within low-level cortices. We therefore contend that the most parsimonious interpretation of the present results is that there are direct, but not forcibly feedforward, interactions between primary auditory and visual cortices of humans.

Multisensory interactions within primary cortices and between rudimentary stimuli require that long-standing notions of cortical organization be revised to include multisensory interactions as a fundamental component of neural organization (e.g. Wallace and others 2004). Here, we show how investigation of BOLD dynamics can address the current gap in knowledge regarding the neurophysiological bases of and brain regions contributing to multisensory interactions.

Supplementary Material

Supplementary material can be found at: http://www.cercor. oxfordjournals.org/.

Notes

Address correspondence to: Roberto Martuzzi, Radiology Service, CHUV, rue du Bugnon 46, 1011 Lausanne, Switzerland. Email: roberto. martuzzi@chuv.ch or to Micah M. Murray, The Functional Electrical Neuroimaging Laboratory, Neuropsychology Division, CHUV, Hôpital Nestlé, 5 Av. Pierre Decker, 1011 Lausanne Switzerland. Email: micah. murray@chuv.ch.

References

Amunts K, Malikovic A, Mohlberg H, Schormann T, Zilles K. 2000. Brodmann's areas 17 and 18 brought into stereotaxic space—where and how variable? Neuroimage 11:66–84.

Beauchamp MS. 2005. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3:93–113.

Beauchamp MS, Argall BD, Bodurka J, Duyn JH. Martin A. 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190-1192.

Bellgowan PS, Saad ZS, Bandettini PA. 2003. Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. Proc Natl Acad Sci USA 100:1415–1419.

Brosch M, Selezneva E, Scheich H. 2005. Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J Neurosci 25:6797–6806.

Calvert GA. 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110-1123.

Cappe C, Barone P. 2005. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886-2902.

Clavagnier S, Falchier A, Kennedy H. 2004. Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect Behav Neurosci 4:117-126.

Falchier A, Clavagnier S, Barone P, Kennedy H. 2002. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749-5759.

Formisano E, Linden DE, Di Salle F, Trojano L, Esposito F, Sack AT, Grossi D, Zanella FE, Goebel R. 2002. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. Neuron 35:185-194.

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE. 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10:77–83.

Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM. 2002. Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol 88:540–543.

Fu KM, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE. 2003. Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23:7510–7515.

Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. 2005. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci 25:5004–5012.

Giard MH, Peronnet F. 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490.

Goense JB, Logothetis NK. 2006. Laminar specificity in monkey V1 using high-resolution SE-fMRI. Magn Reson Imaging 24:381–392.

Gonzalez Andino SL, Murray MM, Foxe JJ, de Peralta Menendez RG. 2005. How single-trial electrical neuroimaging contributes to multisensory research. Exp Brain Res 166:298–304.

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL. 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci 14:6336–6353.

Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ. 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage 15:83–97.

Josephs O, Turner R, Friston K. 1997. Event-related fMRI. Hum Brain Mapp 5:243–248.

Kawashima R, O'Sullivan BT, Roland PE. 1995. Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the "mind's eye". Proc Natl Acad Sci USA 92:5969–5972.

Kayser C, Petkov CI, Augath M, Logothetis NK. 2006. Integration of touch and sound in auditory cortex. Neuron 48:373–384.

Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J Cogn Neurosci 14:420–429.

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.

Logothetis NK. 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971.

Martuzzi R, Murray MM, Maeder PP, Fornari E, Thiran JP, Clarke S, Michel CM, Meuli RA. 2006. Visuo-motor pathways in humans revealed by event-related fMRI. Exp Brain Res 170:472–487.

Miller J. 1982. Divided attention: evidence for coactivation with redundant signals. Cognit Psychol 14:247–279.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115–128.

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ. 2005. Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 15:963–974.

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.

Raab D. 1962. Statistical facilitation of simple reaction times. Trans N Y Acad Sci 24:574–590.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K. 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669–683.

Rockland KS, Ojima H. 2003. Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19–26.

Schroeder CE, Foxe JJ. 2005. Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol 15:454–458.

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC. 2001. Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 85:1322–1327.

Schroeder CE, Mehta A, Givre SJ. 1998. A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575–592.

Schroeder CE, Smiiley J, 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 50:5–18.

Schröger E, Widmann A. 1998. Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology 35:755–759.

Stein BE, Meredith MA. 1993. The merging of the senses. Cambridge, MA: MIT Press.

Tanabe HC, Honda M, Sadato T. 2005. Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J Neurosci 25:6409–6418.

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

Wallace MT, Ramachandran R, Stein BE. 2004. A revised view of sensory cortical parcellation. Proc Natl Acad Sci USA 101:2167–2172.