Exp Brain Res (2007) 179:85–95
DOI 10.1007/s00221-006-0770-6

RESEARCH ARTICLE

Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects

Ryan A. Stevenson · Marisa L. Geoghegan · Thomas W. James

Received: 5 July 2006 / Accepted: 17 October 2006 / Published online: 16 November 2006

Abstract Evidence from neurophysiological studies has shown the superior temporal sulcus (STS) to be a site of audio-visual integration, with neuronal response to audio-visual stimuli exceeding the sum of independent responses to unisensory audio and visual stimuli. However, experimenters have yet to elicit superadditive ($AV > A+V$) blood oxygen-level dependent (BOLD) activation from STS in humans using non-speech objects. Other studies have found integration in the BOLD signal with objects, but only using less stringent criteria to define integration. Using video clips and sounds of hand held tools presented at psychophysical threshold, we were able to elicit BOLD activation to audio-visual objects that surpassed the sum of the BOLD activations to audio and visual stimuli presented independently. Our findings suggest that the properties of the BOLD signal do not limit our ability to detect and define sites of integration using stringent criteria.

Keywords Multisensory · Integration · fMRI · Object recognition · Audio–visual

Introduction

All vertebrates have the ability to extract more than one type of sensory information from the world around them, and have evolved mechanisms for integrating across those different sensory modalities, thereby enhancing their ecological success. Humans are not an exception, and our brains easily integrate multiple sensory inputs into a single consistent perceptual experience under normal circumstances. In the laboratory, however, sensory stimuli can be presented such that two inputs are incongruent and more difficult to integrate. Behavioral studies using this cue-conflict paradigm have uncovered several interesting perceptual phenomena, including "fusion," for which the two sensory inputs are blended into a third distinct percept (McGurk and MacDonald 1976), and "capture", which is when one sensory input dominates the other and it alone is perceived (Mateeff et al. 1985).

Integration can also be studied with congruent sensory inputs by measuring behavioral performance with the combined multisensory inputs and comparing it to performance with each unisensory input. Evidence of enhanced performance with combined stimuli has been found using a variety of behavioral performance measures (for examples, see Hershenson 1962; Morrell 1968). More recently, researchers have become interested in the brain regions and the neural mechanisms that underlie these behavioral enhancements (Meredith and Stein 1983, 1986). Integration has been studied using anatomical and neurophysiological techniques in non-human primates (Benevento et al. 1977; Bruce et al. 1981; Hikosaka et al. 1988; Barraclough et al. 2005), as well as neuroimaging techniques in humans (Calvert et al. 2000, 2001; Beauchamp et al. 2004a, b; Beauchamp 2005).

Classifying an area of cortex or a subcortical structure as a site of multisensory convergence requires operational definitions. Meredith (2002) distinguished between two types of convergence, areal and neuronal.

R. A. Stevenson (✉) · M. L. Geoghegan · T. W. James
Department of Psychological and Brain Sciences,
Indiana University, 1101 East Tenth Street,
Room 293, Bloomington, IN 47405, USA
e-mail: stevenra@indiana.edu

Areal convergence is when two sensory inputs converge on one structure, but the information from those inputs does not interact, or does not interact to a significant degree. Neuronal convergence, or what we will call integration, is when two sensory inputs converge on one structure and the information from those inputs does interact. This definition mirrors what has been found in behavioral studies, that combining sensory inputs produces either a change in the perceptual experience or a change in performance that is different from what would be predicted if the two inputs did not interact (Stein et al. 1988).

The distinction between areal convergence and integration is not difficult to make when recordings from single neurons are available. Many studies utilizing single-unit recordings in non-human primates operationalize multisensory integration using the maximum rule. The maximum rule sets the criterion for integration as the maximum spike count produced by either of the two unisensory inputs ($S_1S_2 > S_1 \cap S_2$). If the multisensory combination stimulus produced a spike count that exceeds the criterion, then the neuron is considered to be integrating the two sensory inputs. Brain structures can be described by the percentage of neurons that are found to exceed the criterion or the percentage of "multisensory" neurons. Although there are solid theoretical grounds for using the maximum rule as a criterion, most studies investigating multisensory regions such as the superior temporal sulcus (STS) and superior colliculus have found augmentation for multisensory stimuli in a subset of cells that not only surpassed the maximum criterion, but that actually surpassed the more stringent superadditive ($S_1S_2 > S_1 + S_2$) criterion (Meredith and Stein 1983, 1986; Barraclough et al. 2005).

Because single-unit recordings are difficult to obtain from humans, to extend the investigation of integration to humans, the field has turned to neuroimaging, particularly functional magnetic resonance imaging (fMRI). We will not provide an exhaustive list here, but will instead focus on studies that investigated audio-visual integration in the STS, a cortical area that is known to contain multisensory neurons in non-human primates. Studies of integration in STS have found enhancements of the blood oxygen-level dependent (BOLD) activation that exceed the superadditive criterion with speech stimuli (Calvert et al. 2000) and with nonsense stimuli (Calvert et al. 2001), and that exceed the maximum or mean rule ($S_1S_2 > (S_1 + S_2)/2$) criteria with non-speech objects (Beauchamp et al. 2004a, b; Beauchamp 2005). All attempts to find superadditivity in BOLD signals using non-speech objects have been unsuccessful, and it has been suggested that the superadditive criterion may be too conservative for non-speech object stimuli (Beauchamp 2005; Laurienti et al. 2005).

One possible explanation for the lack of superadditive findings in BOLD is that fMRI must record from a population of neurons and lacks the capability to record from single neurons. This is a problem because the population within STS is not exclusively multisensory, but also includes unisensory neurons (Benevento et al. 1977; Bruce et al. 1981; Hikosaka et al. 1988; Barraclough et al. 2005). Because fMRI measures a summed signal across all types of neurons, this may allow the activation associated with unisensory neurons to overpower that of multisensory neurons. Also, the enhanced activation with some multisensory neurons may be cancelled out by the attenuated activation with other multisensory neurons (Beauchamp 2005). This concept has previously been modeled in the superior colliculus (Laurienti et al. 2005). In that model, the BOLD activation with audio-visual stimuli was predicted to exceed the maximum rule, but fall short of the superadditivity criterion.

While the neuronal population distribution makes superadditivity difficult to obtain using fMRI, it also makes the distinction between using superadditivity or more liberal criterion (e.g., maximum or mean rule) very important in determining if integration occurs. If activation associated with a population is positive with audio and visual stimuli, where the audio and visual streams do not interact but act independently, then the area would not be integrative. In such an area, the activation with audio-visual stimuli would be the linear sum of the activations with audio and visual stimuli presented independently. Therefore, even if the audio and visual signals do not interact, it would be expected that the activation with audio-visual stimuli would exceed the maximum rule merely by linear summation (Calvert 2001). The same is true for the use of the mean rule, which is by definition equal to or lower than the maximum rule. As such, showing that a population exceeds the maximum rule or the mean rule without achieving superadditivity provides no evidence of integration.

In contrast, finding a population that is superadditive would provide evidence of integration. If the audio and visual signals do not interact, the activation with audio-visual stimuli would be the linear sum of the activations with audio and visual stimuli, and thus would not achieve superadditivity. Conversely, if superadditive activation was found, the audio and visual signals could not be orthogonal, and as such would provide evidence that the two signals were interacting. This shows that the same population-level measurements of BOLD that make it difficult to meet the conservative criteria of superadditivity also make it essential to use superadditivity as opposed to more liberal criteria to identify a multisensory area as a site of integration with BOLD.

A second possible reason that past fMRI studies have been unable to meet superadditivity is a potential vascular ceiling effect in the BOLD signal. If there is strong activation to both audio and visual stimuli when they are presented separately, it is likely that the sum of these two signals will be greater than the level at which the BOLD signal asymptotes, even if the underlying neuronal activity is actually superadditive. An example can be seen in a study by Beauchamp and colleagues (2004a) where BOLD activation with both audio and visual unisensory stimuli was approximately 2% signal change in multisensory regions. With such strong unisensory activations, achieving superadditivity would require 4% or greater BOLD activation, which is highly unlikely in a brain region that is not a primary sensory area. In that case, the experimenters would not have seen superadditivity in the BOLD signal even if multisensory neurons in the area were superadditive, but this would have been due to only the constraints of the BOLD signal itself.

Despite the potential difficulty of meeting the superadditive criterion using fMRI, superadditivity has been found in BOLD activation in two studies to date (Calvert et al. 2000, 2001), one with audio-visual speech stimuli and one with audio-visual nonsense stimuli. Thus, the only recognizable stimuli for which superadditivity has been found are speech stimuli, but speech stimuli have been shown to elicit unique behavioral (McGurk and MacDonald 1976) and neural properties (Narain et al. 2003); therefore it should not be assumed that the results can be generalized to other stimulus classes. This distinction between linguistic and non-linguistic information is common, and has been specifically outlined in reference to integration within STS previously (Calvert 2001). The problem of generalizing to other stimulus classes is further compounded by the fact that fMRI studies of STS using non-speech objects have only reported non-superadditive BOLD activations (Beauchamp et al. 2004a, b; Beauchamp 2005).

In the preceding paragraphs, we described two difficulties with attempting to use a superadditive criterion with BOLD activation measures. In an attempt to overcome these difficulties, we designed an fMRI study using threshold audio and visual stimuli. The first difficulty was that neural activity from unisensory neurons may exert a relatively larger influence over the population neural activity than neural activity from multisensory neurons. In single-unit physiology, integrative response amplification is greater with threshold stimuli than that with supra-threshold stimuli (Meredith and Stein 1983, 1986; Perrault et al. 2005; Stanford et al. 2005) and is called inverse effectiveness. Extrapolating to neural populations suggests that the relative influence of the multisensory neurons on BOLD activation will increase as the discriminability of the stimuli decreases. This would potentially increase the chances of finding superadditive BOLD activation with threshold stimuli compared with supra-threshold stimuli.

The second difficulty was the vascular ceiling effect imposed by the BOLD signal. Low-contrast unisensory visual stimuli produce less BOLD activation relative to high-contrast visual stimuli (Boynton et al. 1996). We expect that using visual and auditory stimuli at psychophysical threshold will produce smaller BOLD activations for both unisensory and multisensory stimuli. If unisensory stimuli produce BOLD activations that are well below vascular ceiling, then it is likely that the multisensory stimuli will also produce an activation that is below the level of the vascular ceiling. Bringing all of the signals below ceiling will allow a better assessment of the superadditive nature of the signals.

Our goal is to obtain superadditive BOLD responses with non-speech object stimuli in a putatively multisensory brain region, the STS. In particular, we will use hand-held tools as stimuli, because these are similar to stimuli used in previous fMRI studies (Beauchamp et al. 2004a, 2005), as well as non-human primate studies (Hikosaka 1988; Barraclough et al. 2005), and as such, will allow us to relate our findings to previous multisensory studies of STS.

Methods

Participants

Participants included eight subjects (five females, mean age = 23). All subjects were right-handed native english speakers. Experiment protocol was approved by the Indiana University Institutional Review Board.

Stimuli

Stimuli consisted of two second digital audio-video recordings of manual tools (e.g., hammer, saw). Scrambled versions of the visual stimuli were created. Each frame was parceled into 20 × 20 pixel squares and exchanging each square with that which had the closest mean luminosity preserving the spatial distribution of luminance. Scrambled versions of the audio stimuli were also created. Each audio stimulus was partitioned and 50% of the waveform was flipped, scrambling the sound but preserving the temporal dynamics of the noise.

Procedures

All stimuli were presented using MATLAB 5.2 (MATHWORKS Inc., Natick, MA) software with the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997), running on a Macintosh computer. Visual stimuli were projected onto a frosted glass screen using a Mitsubishi XL30U projector. Visual stimuli were 200 × 200 pixels and subtended 10.3 × 10.3 of visual angle.

Participants' individual psychophysical thresholds were found while in an MRI simulator designed to mimic the actual fMRI scanner. A three-down-one-up staircase procedure was used to find participants' visual and auditory 79% thresholds independently. Participants were presented with both intact and scrambled stimuli and asked to discriminate between the two. For the visual task, a grid was placed over both intact and scrambled images to prevent discrimination based on artificial contrast gradients produced by scrambling. During the audio task, pre-recorded scanner noise was played at an equal decibel level to the actual scanner. For visual and auditory tasks, dynamic noise (standard deviation = 0.1 and 0.0118, respectively) overlaid the visual and auditory stimuli, and was held constant across staircase levels. Participants completed 200 trials in each modality, and threshold was determined according to the median level of the final 50 trials in each modality. Participants then completed 50 trials at threshold to familiarize themselves with the imaging paradigm.

Each imaging session included two phases: functional localizers and experimental scans. Functional localizers consisted of non-degraded supra-threshold intact stimuli presented in a blocked stimulus design while participants completed a one-back matching task. Runs began with the presentation of a fixation cross for 12 s followed by six blocks of audio-only (A), visual-only (V), or audio–visual (AV) stimuli. AV stimuli were always congruent with audio and video components presented at the same threshold level. Each run included two blocks of each stimulus type, with blocks consisting of eight, two second stimuli presentations, separated by 0.1 s inter-stimuli intervals (ISI). New blocks began every 24 s separated by fixation. Runs ended with 12 s of fixation. Block orders were counterbalanced across runs and participants. Seven participants completed four localizer runs, while one participant completed only two due to time restraints.

During experimental scans, threshold-level intact and scrambled stimuli were presented in a fast event-related design in which participants discriminated between intact and scrambled stimuli. Runs began with the presentation of a fixation cross for 12 s, followed by seven trials of each stimulus type, for a total of 49 trials per run. Stimulus types included intact and scrambled A, V, and AV, as well as fixation. For the seven trials of each stimulus type, four trials were preceded by a two second ISI, two preceded by a four second ISI, and one by a six second ISI. Runs concluded with 12 s of fixation. Trial and ISI orders were counterbalanced across runs and run order was counterbalanced across participants. Seven participants completed six fast event-related runs, and one participant completed only five due to time restraints.

Imaging parameters and analysis

Imaging was carried out using a Siemens Magnetron Trio 3-T whole body scanner, and collected on an eight-channel phased-array head coil. The field of view was 22 × 22 × 9.9 cm, with an in plane resolution of 64 × 64 pixels and 33 axial slices per volume (whole brain), creating a voxel size of 3.44 × 3.44 × 3 mm. Images were collected using a gradient echo EPI (TE = 30 ms, TR = 2000 ms, flip angle = 70°) for BOLD imaging. High-resolution T1-weighted anatomical volumes were acquired using Turbo-flash 3-D (TI = 1,100 ms, TE = 3.93 ms, TR = 14.375 ms, Flip Angle = 12°) with 160 sagittal slices with a thickness of 1 mm and field of view of 224 × 256 (voxel size = 1 × 1 × 1 mm).

Imaging data were pre-processed using Brain Voyager™ 3-D analysis tools. Anatomical volumes were transformed into a common stereotactic space (Talaraich and Tournoux 1988). Functional data were aligned to the first volume of the run closest in time to the anatomical data collection. Each functional run was then aligned to the transformed anatomical volumes, transforming the functional data to a common stereotactic space across participants. Functional data underwent a linear trend removal, 3-D spatial Gaussian filtering (FWHM 6 mm), slice scan time correction, and 3-D motion correction.

Imaging data were analyzed using Brain Voyager™ multi-study general linear model (GLM) procedure. Event-related averages (ERA) were created based on stimulus type for both the localizer and the experimental study using only trials in which subjects responded accurately. A deconvolution analysis was also performed, resulting in the same pattern of activation as the ERA analysis. Hemodynamic peaks were defined as a simple moving average of the time course (2–6 and 6–16 s after stimulus presentation for experimental and localizer scans respectively,). Findings in the left and right hemisphere exhibited the same pattern of activation, with all figures depicting those in the right hemisphere.

Results

Each participants' stimuli contrast level was manipulated in the simulator such that accuracy for both A and V stimuli was 79%. Participants' accuracy was recorded during the experimental scanning in order to ensure threshold levels found in the simulator were valid in the MRI. Participants' mean accuracy for audio threshold stimuli during the experimental scanning was 77.4%, (SEM = 0.04), and for the visual threshold stimuli was 78.9% (SEM = 0.03). These did not significantly differ from 79% target accuracy ($t = 0.43$, $P = 0.68$; $t = 0.01$, $P = 0.99$). Mean accuracy with AV trials was 94.5% (SEM = 0.91) in the simulator and 93.5% (SEM = 1.84) in the MRI, which did not significantly differ ($t = 0.51$, $P = 0.62$). These results verify that participants were performing similarly in the simulator and in the MRI.

Three functional regions of interests (ROI) were defined on an individual subject basis by performing a whole-brain SPM analysis on the localizer runs (in which supra-threshold stimuli were used) and anatomical landmarks (Fig. 1). The visual and auditory ROIs were defined on a map created by contrasting A and V stimuli. This comparison was designed to uncover large areas of cortex that activated more with either A or V stimuli (i.e., unisensory areas). Within the large area that activated more with A than V stimuli, the auditory ROI was defined as a 1,320 mm^3 region of cortex ($x = 55$, $y = -14$, $z = 8$) along the middle of the superior temporal gyrus (STG), which likely corresponds to primary and secondary auditory cortex (Semple and Scott 2003). Within the large area that activated more with V than A stimuli, the visual ROI was defined as a 1,204 mm^3 region of cortex ($x = 42$, $y = -62$, $z = 0$) along the anterior inferior occipital gyrus, which likely corresponds to a portion of the lateral occipital complex (Malach et al. 1995). Multisensory ROIs were defined on a map created by examining the overlap of regions that activated with A and V stimuli. Within the large area that activated with both A and V stimuli, individual multisensory ROIs were defined as a region of cortex on the upper bank of STS (Seltzer and Pandya 1978; Ungerleider and Desimone 1986). These ROIs had a mean volume of 735 mm^3.

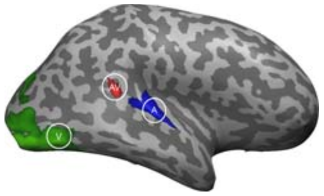

Fig. 1 Unisensory audio, unisensory visual, and multisensory audio-visual ROIs. An example subject's (*LS*) three ROIs, auditory (*A*), visual (*V*), and multisensory (*AV*)

During the experimental runs in which threshold stimuli were presented, BOLD activations with intact A, V, and AV stimuli were extracted from the multisensory ROIs (Fig. 2a). Peak BOLD activation with multisensory AV stimuli was found to be significantly greater than the summed peaks with unisensory A and V stimuli in the right ($t = 2.33$, $P = 0.05$) (Fig. 2b), and left ($t = 2.58$, $P < 0.04$) hemispheres. Activation with AV was greater than the summed activations with A and V for seven out of eight participants. Area under the curve was also analyzed and the result was the same in the right ($t = 2.62$, $P < 0.04$) and left ($t = 2.78$, $P < 0.04$) hemispheres.

In addition to the ROI analysis, a random-effects whole-brain SPM analysis was also conducted on the experimental runs, and a superadditive contrast (A+V > AV) was applied. After correcting for multiple comparisons, there were no clusters that passed statistical threshold. Although finding a cluster in the STS would have shown that the effect described in the ROI analysis was extremely robust, we did not expect this finding, because of the diminished power of whole-brain SPM analyses compared to ROI analyses (Saxe et al. 2006) This finding is similar to previous findings where a whole-brain SPM was also not sensitive enough to detect superadditivity (Beauchamp 2005).

Peak activations with supra-threshold and threshold stimuli were compared within each ROI. Within the audio ROIs, peak BOLD activation with supra-threshold stimuli was found to be significantly greater than with threshold stimuli for A in the right ($t = 16.40$, $P < 0.0001$) and left ($t = 15.31$, $P < 0.0001$) hemispheres, and with AV in the right ($t = 12.99$, $P < 0.0001$) and left ($t = 12.76$, $P < 0.0001$) hemispheres, but not significantly different for V in either the right ($t = 0.13$) or left ($t = 0.59$) hemispheres (Fig. 3a). Within the visual ROIs, peak BOLD activation with supra-threshold stimuli was found to be significantly greater than with threshold stimuli for V in the right ($t = 10.03$, $P < 0.0001$) and left ($t = 11.47$, $P < 0.0001$) hemispheres, and for AV in the right ($t = 14.05$, $P < 0.0001$) and left ($t = 11.85$, $P < 0.0001$) hemispheres, but not significantly different with A in

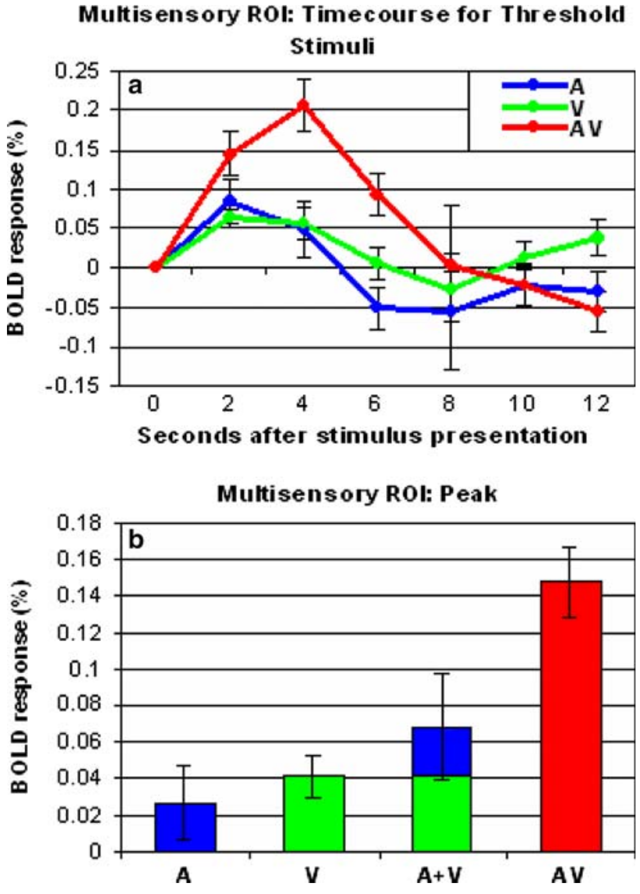

Fig. 2 BOLD activations with unisensory and multisensory stimuli within the right multisensory ROI across subjects. **a** Average time courses across participants ($n = 8$) of BOLD activation within the multisensory ROI, depicting activations with unisensory auditory (*blue*), unisensory visual (*green*), and multisensory audio-visual (*red*) stimuli. **b** Peak BOLD activations with unisensory auditory (*blue*), unisensory visual (*green*), and multisensory audio-visual (*red*) stimuli, as well as the summed peak of BOLD activations with unisensory audio and visual (*blue* and *green stacked*) stimuli for comparison. Error bars represent SEM

either the right ($t = 1.70$) or left ($t = 1.22$) hemispheres (Fig. 3b). Within the multisensory ROIs, peak BOLD activation with supra-threshold stimuli was found to be significantly greater than to threshold stimuli with A in the right ($t = 3.82$, $P < 0.005$) and left ($t = 5.60$, $P < 0.0005$), and with V in the right ($t = 8.93$, $P < 0.0001$) and left ($t = 9.20$, $P < 0.001$), but not significantly different with AV in the right ($t = 2.01$) or left ($t = 2.01$) hemisphere (Fig. 3c).

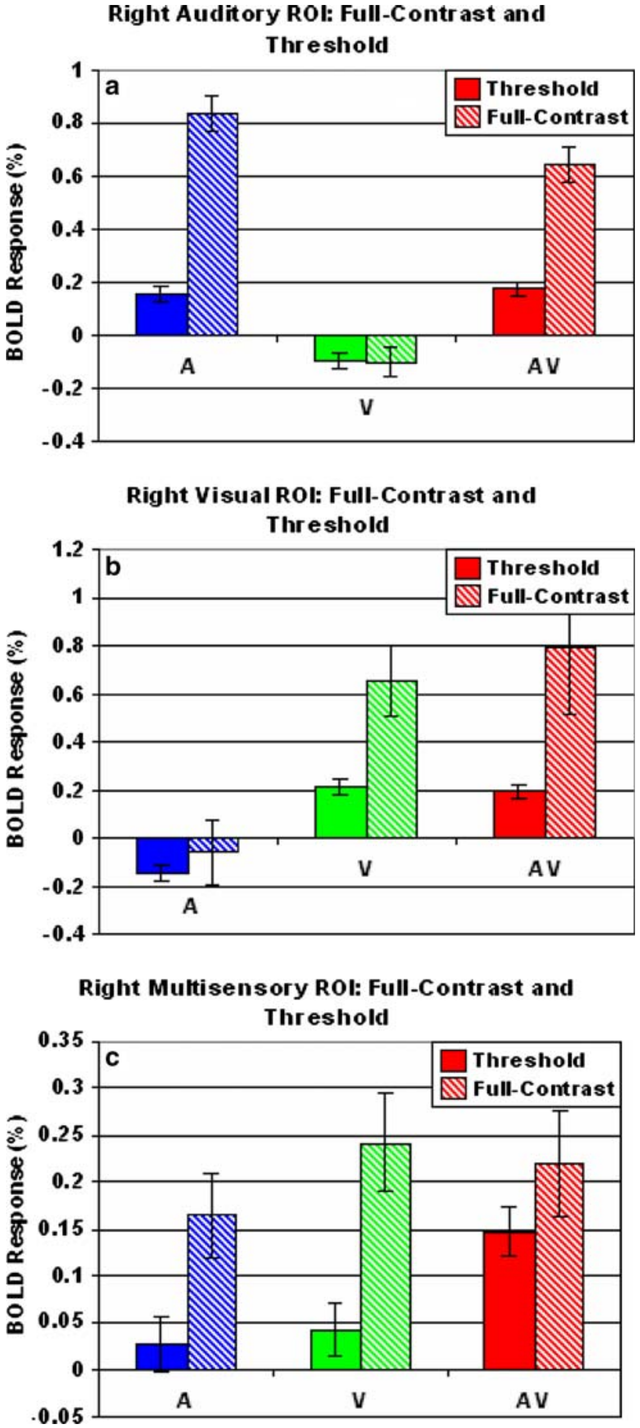

Fig. 3 Peak BOLD activations with threshold and supra-threshold stimuli within each right ROI. Peak BOLD activations with unisensory audio (*blue*), unisensory visual (*green*), and multisensory audio-visual stimuli (*red*). *Solid bars* indicate activations with threshold stimuli and *hashed bars* indicate activations with supra-threshold stimuli. BOLD activations are shown within the auditory (**a**), visual (**b**), and multisensory (**c**) ROIs. Error bars represent SEM

In addition, BOLD activation for intact and scrambled stimuli with A, V, and AV stimuli presented during the experimental runs (in which threshold stimuli were presented) were extracted from the ROIs and compared. Within the audio ROIs, peak BOLD activation with intact stimuli was significantly greater than with scrambled stimuli with A in the right ($t = 2.97$, $P < 0.03$) and left ($t = 2.55$, $P < 0.04$) hemispheres, and with AV in the right ($t = 3.00$, $P < 0.03$) and left ($t = 3.01$, $P < 0.03$) hemispheres, but not significantly different with V in either the right ($t = 0.15$) or left ($t = 0.32$) hemispheres (Fig. 4a). Within the visual ROIs, peak BOLD activation with intact stimuli was found to be significantly greater than with scrambled stimuli with V in the right ($t = 2.85$, $P < 0.03$) and left ($t = 2.41$, $P < 0.05$), and with AV in the right ($t = 3.00$, $P < 0.03$) and left ($t = 3.67$, $P < 0.01$) hemispheres, but not significantly different with A in either the right ($t = 0.11$) or left ($t = 0.25$) hemispheres (Fig. 4b). Within the multisensory ROIs, peak BOLD activation with intact stimuli was not found to be significantly greater than with scrambled stimuli with A with A in either the right ($t = 0.57$) or left ($t = 0.81$) hemispheres, nor with V in either the right ($t = 1.98$, $P < 0.1$) or left ($t = 0.56$) hemispheres, but was significantly greater with AV in the right ($t = 2.35$, $P = 0.04$) but not the left ($t = 0.92$) hemisphere (Fig. 4c).

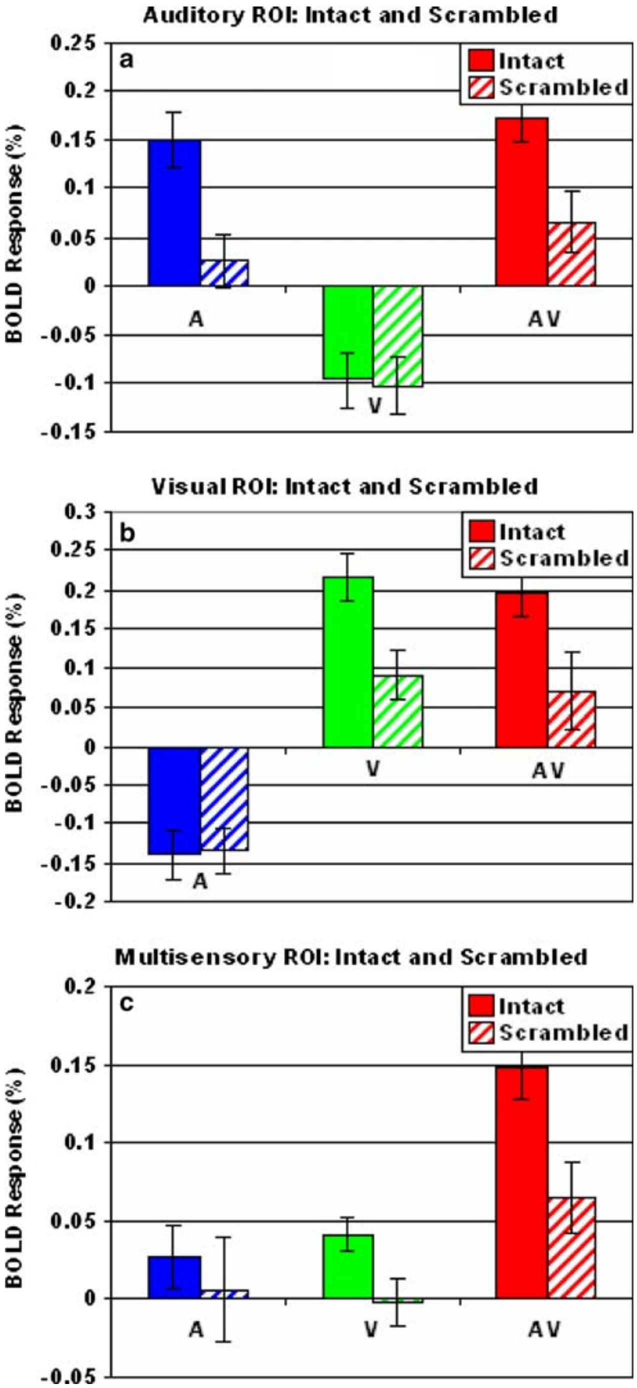

Fig. 4 Peak BOLD activations with intact and supra-threshold stimuli within each right ROI. Peak BOLD activations with unisensory audio (*blue*), unisensory visual (*green*), and multisensory audio-visual stimuli (*red*). *Solid bars* indicate activations with intact stimuli and *hashed bars* indicate activations with scrambled stimuli. BOLD activations are shown within the auditory (**a**), visual (**b**), and multisensory (**c**) ROIs. Error bars represent SEM

Discussion

Comparing unisensory and multisensory stimuli

Using threshold non-speech stimuli, we were able to show the same superadditive increase in BOLD activation in STS that has been previously reported for speech stimuli (Calvert et al. 2000). To our knowledge, this is the first experiment using non-speech objects to find superadditive BOLD activation in any brain region. This pattern of superadditivity was found bilaterally, in contrast to findings with speech stimuli and nonsense stimuli, which are lateralized to the left hemisphere (Calvert et al. 2000, 2001). Other neuroimaging studies have claimed to find integrative BOLD activation in STS with non-speech objects, but these claims have been based on more liberal criteria, such as the maximum rule (Beauchamp et al. 2004a) or the mean rule (Beauchamp et al. 2004b). Our findings show that superadditivity, which is a more conservative criterion, can be used successfully in fMRI studies with other classes of audio-visual multisensory stimuli than speech.

Although we are only beginning to understand the neural mechanisms involved in integration, it is clear that the maximum rule criterion has significance when applied to measures of single-unit activity. When a multisensory stimulus causes a neuron to fire more than either of the unisensory inputs, information from the two inputs must be interacting (Meredith 2002). But, when studying integration in populations of neurons, such as with fMRI, different assumptions must be made. When two sensory inputs converge on one brain region, they do not necessarily have to interact. This has been described elsewhere as areal convergence (Meredith 2002). Different groups of neurons may be present in a brain region that may be isolated from each other, receiving separate inputs and sending separate outputs, without their signals ever interacting. When this idea is applied to the BOLD signal, which can only measure populations of neurons, the use of more liberal criteria, such as the maximum or mean rule, becomes tenuous. There is evidence that the STS shows multisensory enhancement with non-speech audio-visual stimuli at a level above the maximum and mean rule criterion (Beauchamp et al. 2004a, 2004b; Beauchamp 2005). This level of activation, however, can be explained without any interaction between the auditory and visual inputs. If STS has one group of audio neurons that receive audio-only inputs, and send outputs to audio-only areas, and one group of visual neurons that receive visual-only inputs and send output to visual-only areas, then the activation with audio-visual stimuli would exceed the maximum rule. Thus, the maximum rule criterion cannot distinguish between areal convergence and true integration.

The more conservative criterion of superadditivity, which surpasses the level of enhancement that can be explained by two non-interactive information streams, can distinguish between areal convergence and integration. In the case of areal convergence, the BOLD signal would be the linear sum of the activations with unisensory stimuli. Superadditivity by definition exceeds that sum; therefore, finding enhancement that exceeds the superadditivity criterion implies integration of the two sensory inputs.

Superadditivity has been clearly shown in single-unit recordings of neurons in STS (Barraclough et al. 2005), leaving little doubt that STS is a site of integration. Superadditive BOLD activation, however, has been difficult to achieve. This could be due to vascular ceiling effects in the BOLD signal, to the heterogeneity of neurons making up the populations, or because BOLD activation does not reflect neural spiking. By using threshold stimuli, we were able to show superadditive BOLD activation with non-speech objects, suggesting that BOLD imaging can be used to detect superadditive activation. Furthermore, our findings show mean activation with AV stimuli that was 118% greater than the sum of the individual mean activations with A and V, a result that is well within the range of enhancements found in individual neurons (Barraclough et al. 2005). Together, these results suggest that the lack of BOLD effects in previous studies was not due to complete insensitivity of BOLD measurements to superadditive patterns of neural activity, but rather was due to asymptotic BOLD activation.

It should be noted here that the neural properties underlying the BOLD signal are still not completely clear. Of particular interest here is the discovery that the spiking output of neurons is not the most predictive measure of BOLD activation (for review, see Heeger et al. (2000); Logothetis et al. (2001); Atwell and Iadecola (2002); Logothetis (2002, 2003); Logothetis and Wandell (2004)). In fact, it appears that the total synaptic activity of a neuron is more predictive of BOLD activation. Thus, care should be taken when attempting to make predictions about BOLD activation based on single-unit recordings and vise versa. The ambiguity of the relationship between neural activity and BOLD response poses another problem for research on multisensory integration: the use of superadditivity as a criterion for multisensory integration in single-units is quantitatively sound, because spike counts are measured on a ratio scale, that is, they have a defined zero point. Zero BOLD activation, on the other hand, is not necessarily related to zero neural activity or even to spontaneous firing at rest, nor is a resting condition necessarily relateable to zero BOLD activity (Stark and Squire 2001; Binder et al. 1999). Thus, until the relationship between neural activity and BOLD activation has been quantified, particularly during the resting state, reports of superadditivity should be interpreted with caution. To facilitate interpretation of our results, we have presented our data as percent signal change values, a scale that is fairly universal for neuroimaging data. Furthermore, to evince enough statistical power from our design, we used a rapid event-related design with a distribution of ISIs between 2 and 14 s, exponentially distributed (see Methods). This type of distribution of ISIs has been shown in simulations to provide excellent estimates of BOLD activation (Birn et al. 2002; Serences 2004), whether using event-related averaging or deconvolution analysis techniques. For all of the results presented here, we used the averaging technique, but using deconvolution produced the same pattern in terms of statistical significance.

Comparing supra-threshold and threshold stimuli

We compared BOLD activations with intact threshold stimuli to those with the intact supra-threshold stimuli. Decreasing the contrast or intensity of the stimuli decreased the BOLD activation in regions that were sensitive to that sensory modality. For instance, in the auditory ROI, the activation with threshold A stimuli was less than with supra-threshold A stimuli, whereas there was no change in activation with V stimuli. In the auditory ROI, the AV stimuli followed the same pattern as the A stimuli, presumably because the activation in this ROI was driven mainly by the auditory signal. Likewise, in the visual ROI, there was a significant decrease in activation with the threshold stimuli with V and AV, but no change in activation with A.

The multisensory ROI showed a different pattern. BOLD activations were significantly reduced with threshold A, V, and V stimuli, but were reduced less with threshold AV stimuli, compared with supra-threshold counterparts.

It is likely that our ability to detect the superadditive BOLD activation enhancement with non-speech objects is related to the decreased activation produced with threshold stimuli. The reduction in BOLD activation with unisensory stimuli in the multisensory ROI lessened the chance that BOLD activation with multisensory stimuli would reach its vascular ceiling. The superadditivity criterion value for threshold stimuli (found by summing the individual activations with the unisensory stimuli) was 0.07% signal change, which was substantially lower than 0.24%, the level at which BOLD activation asymptoted with our supra-threshold stimuli. These findings suggest that the reason for other failures to find superaddivity may be related to the vascular ceiling effect in the BOLD signal.

Overcoming the population distribution obstacle requires a quieting of unisensory activation while not significantly diminishing multisensory activation. Using threshold stimuli reduced activation with unisensory stimuli to a much greater extent than activation with AV stimuli within the multisensory ROI. What causes this larger activation reduction with specifically unisensory stimuli when using threshold stimuli is currently unknown. A number of studies have shown that multisensory neurons respond with a greater superadditive response enhancement with threshold stimuli than with supra-threshold stimuli due to inverse effectiveness (Meredith and Stein 1983, 1986; Perrault et al. 2005; Stanford et al. 2005). Therefore, using threshold instead of supra-threshold stimuli may have increased the relative contribution of multisensory neurons to BOLD activation. Further investigation of the function describing the interaction of stimuli contrast and BOLD activation in STS to both unisensory and multisensory stimuli would be a logical step to take in understanding the multisensory interaction within STS.

Researchers have also suggested that the inability to find superadditivity in BOLD activation with non-speech objects may have been due to attenuation neurons. Attenuation neurons respond strongly to one specific modality when presented in isolation, but only weakly to multisensory stimuli and other unisensory modalities (Barraclough 2005). As such, researchers have previously theorized that these neuronal response properties essentially cancel out the activation associated with superadditive neurons (Beauchamp 2005). However, in our data, there is a greater reduction of BOLD activation from supra-threshold to threshold with unisensory stimuli in the right (A = 84%, V = 83%) and left (A = 91%, V = 93%) hemispheres than with multisensory stimuli in both the right (AV = 32%) and left (AV = 39%) hemispheres (Fig. 3c). This pattern of BOLD activation, while not directly measuring neuronal output, suggests that the activity of attenuation neurons may experience a floor effect with threshold stimuli, which would lessen their effect on BOLD activation. Thus, using threshold stimuli may increase the relative contribution of activity from superadditive neurons to BOLD activation, and at the same time diminish the relative contribution of activity from attenuation neurons.

Comparing intact and scrambled stimuli

The comparison of intact to scrambled stimuli was made to determine if activation in STS is sensitive to recognizable objects, as has been previously suggested (Calvert et al. 2001; Beauchamp 2004b; Amedi et al. 2005). Differences in BOLD activation of scrambled stimuli and intact stimuli within audio and visual ROIs showed that activations with scrambled stimuli were significantly less than those to intact stimuli, but only when the stimuli contained the area's preferred modality. In the auditory ROI, there was a significant decrease in activation with the scrambled stimuli with A and AV stimuli, but no change in activation with V stimuli. Likewise, within the visual ROI, there was a significant decrease in activation with the scrambled stimuli with V and AV stimuli, but no change in activation with A stimuli. Within the multisensory ROI, there was a general overarching pattern of signal reduction with scrambled stimuli. This trend of reduction in STS with scrambled stimuli as compared to intact stimuli is further evidence that STS is more sensitive to identifiable objects (Beauchamp 2004b; Amedi et al. 2005).

While there is evidence for a general reduction in BOLD activation to scrambled stimuli, the superadditive pattern relating activations with unisensory and multisensory stimuli remained unchanged. The stimuli were scrambled in such a way that while object identification was affected, the temporal and spatial properties remained intact and congruent, stimulus properties that have been shown to be definitive factors in superadditive neural responses of individual cells (Hershenson 1962; Morrell 1968; Meredith and Stein 1986, 1996; Meredith et al. 1987) and BOLD activations (Calvert et al. 2000) with multisensory stimuli. The consistency in the pattern of BOLD activation to unisensory and multisensory stimuli, even when there is not an identifiable object, provides evidence that this pattern is not due to object identification, but may be in response to spatially and temporally congruent AV stimuli. The retention of this pattern of superadditivity in response to scrambled stimuli further supports the idea that these findings can be generalized to other stimuli.

Conclusion

Our findings demonstrate that STS is a site of multisensory integration for non-speech objects. Furthermore, our results demonstrate that superadditivity may not be an overly stringent criterion for use with fMRI and can be elicited in known integrative regions such as STS.

Acknowledgments This research was supported in part by the Indiana METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. Thanks to Karin James and Laurel Stevenson, as well as James Townsend, Ami Eidels, and the Indiana University Neuroimaging Group for their insights on this work and manuscript.

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

Attwell D, Iadecola C (2002) The neural basis of functional brain imaging signals. Trends Neurosci 25:621–625

Barraclough NE, Xiao D, Baker CI, Oram MW, Perret DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377–391

Beauchamp MS (2005) Statistical criteria in fMRI studies od multisensory integration. Neuroinformatics 3:93–113

Beauchamp MS, Argall BD, Bordurka J, Duyn JH, Martin A (2004a) Unraveling multisenory integration: patchy organization within human STS multisensory cortex. Nat Neurosci 7:1190–1192

Beauchamp MS, Lee KE, Argall BD, Martin A (2004b) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823

Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol 57:849–872

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999) Conceptual processing during the conscious resting state: a functional fMRI study. J Cogn Neurosci 11:80–93

Birn RM, Cox RW, Bandettini PA (2002) Detection versus estimation in event-related fMRI: choosing the optimal stimulus timing. Neuroimage 15:252–264

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 26:369–384

Calvet GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438

Heeger DJ, Huk AC, Geisler WS, Albrecht AG (2000) Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? Nat Neurosci 3:631–633

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289–293

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

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 (2002) The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci 357:1003–1037

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

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157

Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Annu Rev Physiol 66:735–769

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139

Mateeff S, Hohnsbein J, Noack T (1985) Dynamic visual capture: apparent auditory motion induced by a moving visual target. Perception 14:721–727

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

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

Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in the cat superior colliculus neurons I: temporal factors. J Neurosci 7:3215–3229

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389–391

Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res 365:350–354

Meredith MA, Stein BE (1996) Spatial determinates of multisensory integration in cat superior colliculus. J Neurophysiol 75:1843–1857

Morrell LK (1968) Temporal characteristics of sensory interaction in choice reaction times. J Exp Psychol 77:14–18

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Mathews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362–1368

Pelli DG (1997) The video toolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442

Perrault TJ Jr, Vaughn JW, Stein BE, Wallace MT (2005) Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. J Neurophysiol 93:2575–2586

Saxe R, Brett M, Kenwisher N (2006) Divide and conquer: a defense of functional localizers. Neuroimage 30:1088–1096

Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex. Brain Res 149:1–24

Semple MN, Scott SK (2003) Cortical mechanisms in hearing. Curr Opin Neurobiol 13:167–173

Serences JT (2004) A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. Neuroimage 21:1690–1700

Stanford TR, Quessy S, Stein BE (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. J Neurosci 25:6499–6508

Stark CE, Squire LR (2001) When zero is not zero: the problem of ambiguous baseline conditions in fMRI. Proc Natl Acad Sci USA 98:12760–12766

Stein BE, Huneycutt WS, Meredith MA (1988) Neurons and behavior: the same rules of multisensory integration apply. Brain Res 448:355–358

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Thieme Medical Publishers, New York

Ungeleider LG, Desimone R (1986) Projections to the superior temporal sulcus from the central and peripheral field representations of V1 and V2. J Comp Neurol 248:147–163